

80 03581
v.1

SMALL AREA POPULATION
PROJECTIONS FOR
HEALTH PLANNING

The Association of Bay Area
Governments
Berkeley, California

June 1977

The production of this book was supported by a subcontract with the Health Systems Management Corporation entitled the "Joint Health Planning Project." The Joint Health Planning Program was funded by the Bureau of Health Planning and Resources Development, Health Resources Administration, Department of Health, Education and Welfare under Contract Number HRA 230-75-0069.

THE UNIVERSITY OF CHICAGO
LIBRARY
5401 S. UNIVERSITY AVE.
CHICAGO, ILL. 60637

THE UNIVERSITY OF CHICAGO
LIBRARY
5401 S. UNIVERSITY AVE.
CHICAGO, ILL. 60637

June 1977

The University of Chicago will accept for deposit
any book or manuscript which is published in the
United States or elsewhere, and which is not
already in the possession of the University.
Books and manuscripts deposited with the
University are subject to the provisions of the
Copyright Act of 1909, and the provisions of the
Copyright Act of 1976, and the provisions of the
Copyright Act of 1992.

ACKNOWLEDGEMENTS

Many persons played contributing roles during the two year project. Nancy Gibbons Shyer contributed substantially to the demographic research. She had the major responsibility for preparing the regional population projections. Adela Castro provided computer programming, table and graphics development and research support. A special thanks to them, for the project would have suffered without their energy.

Stan Hoffman provided ideas and thoughtful suggestions as the project developed. Carl Gerlach and Ann Midkiff, both of HSMC, were always available to test out ideas and new materials. Stan often acted as the catalyst for all of us to keep things moving.

Finally, the cover design was by Pat Wong-Yung. The maps were prepared by Karen Graser. Jennifer Wolch was a sharp editor.

Phil Lankford
Program Manager

TABLE OF CONTENTS

	Page
I. Conceptual Overview	1
Introduction	1
The Regional Concept	1
The Demographic Model	2
Subarea Population Projections	2
Social Area Analysis	4
Subarea Definition	4
The Allocation Procedure	4
Summary of Major Findings	5
Conclusions	6
II. The Users Manual for Health Planners	8
1. The Regional Approach	8
2. Developing Regional Population Projections	9
An Overview of the Regional Population Projections Model	9
Modular Design	10
Uses of APPLE	10
CORE: The Cohort-Survival Model	12
Fertility	21
Mortality	26
Migration	32
Household Headship Rates	34
Labor Force Participation Rates	37
Other Parameters	38
Range of Regional Population Projection Assumptions	49
APPLE Sensitivity Analysis	67
The Role of an Advisory Committee	71
Producing the Final Regional Projections	71
Summary	83
3. Social Area Analysis	84
The Factor Analysis Method	84
Factor Analysis Computer Programs	90
Data	90
Selection of Variables	90
Data Processing	94
Data Processor Design	94
Factor Analysis Results	95
Suggested Bibliography	124

TABLE OF CONTENTS (Continued)

	Page
4. Community Definition	126
Tract Similarity Measurement	126
Single Linkage Algorithms	126
Neighborhood Limited Approach	127
Neighbor Definition	128
Contiguity Matrix	128
5. Small Area Population Allocations	156
The Allocation Process	156
Results and Validation	160
6. Summary and Shortcuts	161

LIST OF TABLES

Table	Page
1. Projections and Estimates of Fertility Rates	22
2. Sample Calculation of Age Specific Fertility Rates	23
3. Sample Life Table	28
4. Forward Survival Method	33
5. Calculation of Household Headship Rates	36
6. Calculation of Labor Force Participation Rates	39
7. Fertility Assumptions	54
8. Proportion of Births Born by Cohort	55
9. Regional Mortality Rates and Projections	56
10. Labor Force Participation Rates	58
11. Number of Households	62
12. Base Year Household Headship Rates	64
13. Projected Changes in Household Headship Rates	65
14. Summary of Final Projection Assumptions	74
15. Age and Sex Distribution of Regional Population	78
16. Factor Matrix	87
17. Rotated Factor Loadings	89
18. Social Area Analysis Variables	91
19. Second Pass Univariate Statistics	97
20. First Pass - 14 Factor Structure	99
21. First Pass - 5 Factor Solution	100
22. First Pass - 14 Factor Solution	103

LIST OF TABLES (Continued)

Table	Page
23. Second Pass Solutions	108
24. Second Pass - 6 Factor Solution	109
25. Second Pass - 7 Factor Solution	111
26. Second Pass - 8 Factor Solution	114
27. Second Pass - 9 Factor Solution	116
28. Second Pass - 10 Factor Solution	118
29. Third Pass	120
30. Neighbor Distribution	130
31. Grouping of 440 Zones	131
32. Sample Social Area Profile	141
33. Sample Social Area Age/Sex Structure	158
34. Sample Social Area Population Projection	159

LIST OF FIGURES

Figure	Page
1. Cohort-Survival Method of Projecting	3
2. Overview of the APPLE System	11
3. CORE Flow Diagram	13
4. Total Fertility Rate, 1835 to 1975	50
5. Total Fertility Rate, 1950 to 2000	50
6. Total Fertility Rate, 1960 to 1975	51
7. Household Size	66
8. Sensitivity Analyses	72
9. Average Annual Births, Deaths, Net Migrations, 1975-2000	75
10. Projected Changes in the Proportions of Youth and Aged, 1990,2000 .	76
11. Reduction of Data Matrix to a Matrix of Weighted Scores	85
12. Graphical Representation of Table 16 and Rotation	87
13. Rooks Case	129

LIST OF MAPS

Seven Maps follow page 123.

Map

1. Lifestyle
2. Stage in Life Cycle
3. Black Subpopulation
4. Women in the Labor Force
5. Hispanic Subpopulation
6. Asian Subpopulation
7. Group Quarters

Map 8, Social Area Delineation, follows page 139.

PART 1
CONCEPTUAL OVERVIEW

I. CONCEPTUAL OVERVIEW

Introduction

The status and improvement of health is a current issue of national, state, and local importance. As we ask more from and about the health care industry, the need for analysis and projections of health care resources demand increases, particularly at the small area level where the link between provider and consumer is critical.

The age and sex composition of the population using a health care resource is the major determinant of its total demand function. Aggregate total population figures for large areas become useless in answering the complex supply and location issues of today. Health planners need population projections by age and sex for small areas. Useful methods for the definition of small areas appropriate to health care analysis and for producing detailed small area population projections are presented in this document.

The document is organized in three sections and bound in two volumes. The first section is a conceptual overview of the entire set of methods used. It shows how the different methods relate to each other within a general framework. The second section, bound with the first, is the "how to do it" manual. It is intended for health planners with a technical background. The final section, bound separately, supplies the computer programs needed to implement the methods.

The approach to small area population projections draws upon modern geographical and sociological research. Small areas are viewed as well-defined parts of a region. Population projections are first made for the region, totals assigned to each subarea, and then broken into age/sex groups based on the subarea's characteristics.

The Regional Concept

An integrated approach is necessary for small area population work. Many problems arise in population studies when the study area is "open", i.e., with much of its population commuting each day from another area. Another complication when projecting population over several decades is migration. Migration information is generally not available for small areas. From the health resources view there is always much overlap in the "market" or service areas of the health provider institutions. To correctly study the relationships between supply and demand the area must be large enough to contain all interactions. By adopting a regional perspective all these problems can be solved.

The region should be a functional unit, large enough to encompass all of the economic and social activities of the population within it. The boundary of the region can generally best be defined as containing all of the daily commuting patterns. Such a daily commuting field then contains all of the interactions (social, economic, cultural) and institutional settings

(provider service areas). All of the concepts and methods below work within this regional context.

The Demographic Model

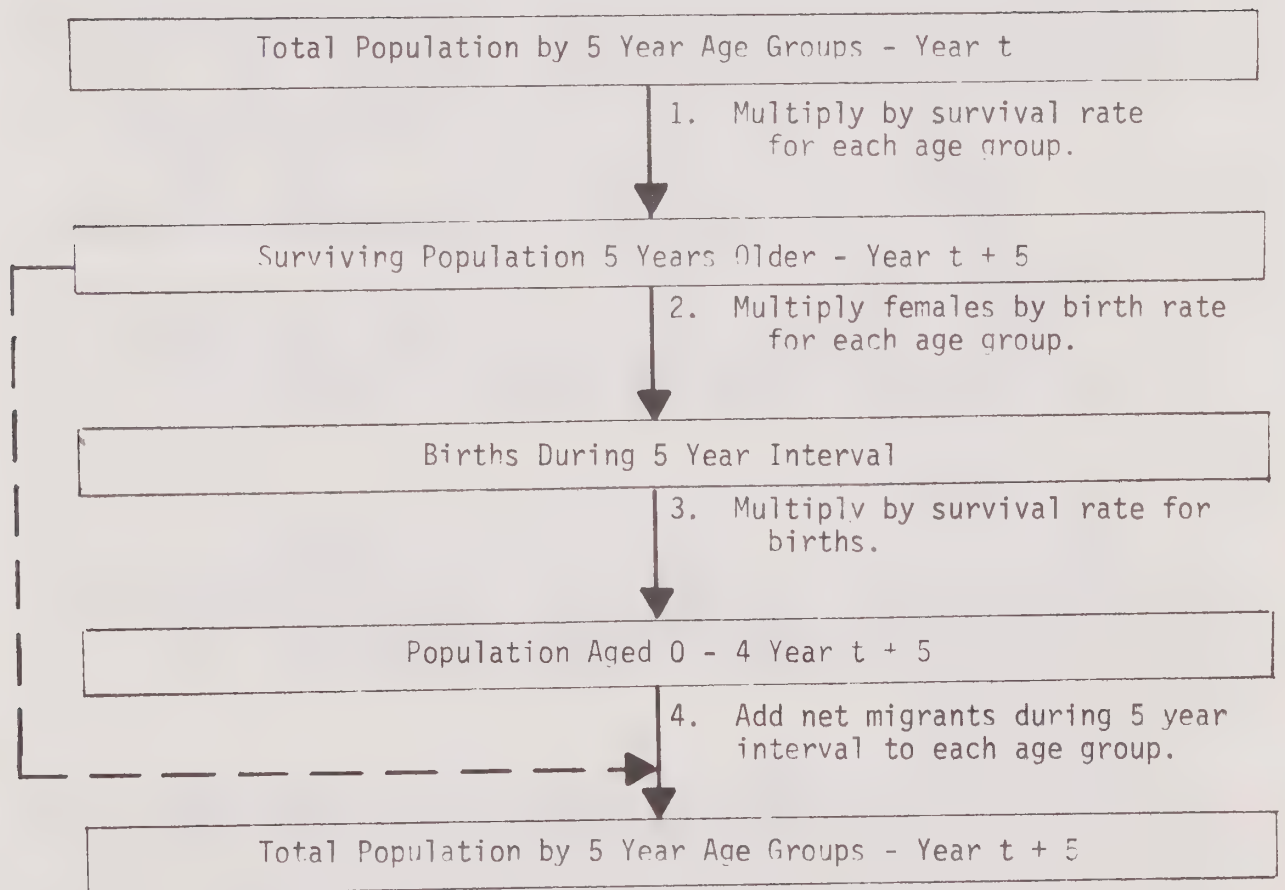
The health planner needs detailed population projections. The population must be broken into groups based on age and sex, called cohorts. Age is divided into five year intervals. The age-sex cohorts, during a five year projection period, are subject to the three major components of population change: (1) decreases due to death, (2) increases resulting from births, and (3) increases or decreases resulting from in- and out-migration. The only demographic method which considers these three types of changes at the level of detail needed is called the cohort-survival method. As illustrated in Figure 1, the basic procedure begins with a base year population (e.g., 1970) divided into five-year age-sex groups (males 0-4, males 5-9, etc.).

In step 1, the population in each cohort is moved five years forward in time by multiplying the number of people in the age group by the probability of surviving five years (e.g., population aged 15-19 in 1970 is "survived" to ages 20-24 in 1975). The step determines the number of survivors in each age group five years in the future for ages 5 and above. The next step (step 2 in Figure 1) is to calculate the number of births during the five-year period. Births are derived by multiplying the number of women in each five-year age group by the probability of a woman in that age group bearing a child. In step 3, the births are multiplied by the probability of surviving to the end of the interval (1975) in order to derive the number of children aged 0-4 at the end of the interval. Finally, in step 4, the gains or losses due to migration are added to each cohort. The result is the complete age-sex distribution at the end of a five-year cycle. The computations can be repeated for as long a projection period as desired. The assumptions about future trends in fertility, mortality, migration and other demographic parameters are inputs to the model. Since these trends, particularly migration, can only be systematically studied at the regional level the cohort-survival model is operated for the entire region.

Subarea Population Projections

Population projections for small subareas have been a major problem in planning. The failure of past approaches has been the treatment of each subarea as an arbitrary unit rather than as a part of an integrated whole. The methods used in this approach consider the subarea projections primarily as a problem in allocation of future regional population to rationally defined subareas. Before the approach can be used to allocate regional population to small areas we must know the social, economic, and demographic characteristics and their spatial distribution within the region which assist in formulation of a rational subarea description. The technique which allows us to understand the social structure of a region and the spatial pattern of the social fabric is termed social area analysis. The results of the analysis are used to define social areas, or communities, and are the basis of subarea projections.

FIGURE 1
COHORT-SURVIVAL METHOD OF PROJECTING



Social Area Analysis

Cities display regularities and stability in the spatial distribution of their social and demographic characteristics over time. Social area analysis, also termed factorial ecology when refined, examines the spatial pattern of population characteristics by seeking the basic dimensions needed to describe the social spatial structure. Normally a large number of socioeconomic, housing, ethnic, labor force characteristics, and other variables are reduced to their latent dimensions by a statistical method called factor analysis.

During the past thirty years enough studies have been completed to recognize three basic types of dimensions for characterizing populations:

1. Socioeconomic status of individuals, involving educational background, type of employment, level of income, quality of residence, etc. This factor varies by sectors, with lower status sectors following jobs and higher-status sectors seeking out superior residential amenities.
2. Stage of families in the life cycle, which determines family size, type of housing and location of residence selected, attitudes to density, use of personal automobile, etc.
3. Segregation, the restriction of certain groups to particular communities within the city.

As a function of the factorial approach, the three dimensions are independent and additive. For each unit of analysis, normally a census tract, we define its position in physical space as well as social space with its statistical "score" or position on each basic dimension.

Subarea Definition

A grouping or clustering analysis is used to join census tracts with similar characteristics. A geographic constraint is applied to the analysis to ensure that similar tracts are spatially adjacent before they are joined into a social area. As the tracts are grouped, social areas or communities are developed. Since the areas are homogeneous with respect to social status, stage in family life cycle and ethnic composition, they are an excellent basis for small area population projections and health planning interpretation.

The Allocation Procedure

The homogeneous nature of the social areas is the major factor in the allocation of population by age and sex. As several studies have shown, a social area's general age/sex structure is relatively stable over a fifteen to twenty-year period. A secondary factor is the general pattern of population change exhibited in the regional demographic model. The subareas share of the regional population total is derived from a land

use model. The social areas' total population is allocated to the age/sex cohorts using the base year (1970) characteristics adjusted by the general regional demographic trends. For instance, as a result of the aging of the post World War II baby boom, and the recent sharp declines in fertility, there is a shift towards an older age/sex distribution.

A key input to the allocations is the projected subregional total population for each community. This value is an output of the projected land use model, PLUM. The PLUM model, developed over several years, incorporates trends in commuting patterns, employment opportunities and location, shifts in land available for development and a wide range of other parameters that describe land use. For each five-year period PLUM produces a projection of total population for small subareas. These subareas are aggregated to the community definition employed by health planners. The grand total of the subarea populations equals the projections from the regional demographic model.

Summary of Major Findings

The nine counties of the San Francisco Bay Area are an excellent laboratory to test health planning methods. The region contains a diverse ethnic, social, and economic mixture within urban and rural communities. The methods developed and tested under the project have general applicability and transferability. Some of the specific findings are listed below.

- A. The nation, California, and the region are all experiencing unprecedented declines in fertility. Not only are we below the replacement level, but we are at a level below that of the Great Depression (1930-38). Changes in the assumed future fertility level can make significant changes in the composition and size of future regional populations.
- B. Migration is the most difficult demographic parameter to estimate and future levels can potentially generate the greatest changes in regional population. Migration is probably the single most significant variable in population projections.
- C. The region's population can be best characterized by seven specific dimensions, which are also the basis for forming the subareas. The seven key descriptions of the social fabric of the Bay Area were:
 1. Lifestyle - a choice of suburban or urban existence.
 2. Stage in Life Cycle - how people locate in social and physical space to meet their housing demands through the changes life brings.
 3. Black Subpopulation - also describes families in economic distress and working in the low paying services sector.

4. Women in the Labor Force - also indicates a high employee to resident ratio.
 5. Hispanic-Heritage Subpopulation - also describes inadequate housing, poor education and many children.
 6. Asian Subpopulation - dominated by Chinese, but other minority groups, such as Filipino, are described.
 7. Group Quarters Population - indicates special institutional housing arrangements.
- D. Ninety-six homogeneous subareas were defined for the region, ranging in population from 4,000 to 180,000.
 - E. By the nature of their homogeneity, the social areas provide an early, reliable method to convert the ZIP Code-based health data into census tract-based areas.
 - F. The 1975 small area projections were compared to the 1975 census for the parts of the region that participated in the mid-decade effort. The models did quite well over this five-year period. Further validation would be appropriate when the 1980 Censuses are collected.

Conclusions

The regional approach is successful compared to other small area population projection efforts. Originally the techniques were aimed at urban populations over 250,000. However, they are generalized to apply to lesser populated regions. Short-cut methods are suggested to minimize staff and computer costs. The project has developed and tested a methodology and implemented the necessary computer software to: (1) produce detailed population projections by age and sex for small subareas, and (2) establish a definition of subareas suitable to health planning analysis. However, the approach taken has gone beyond merely packaging techniques. It has developed an integrated interpretive framework for planning analysis of spatial problems found in the health care delivery system. The social area analysis allows the health planner to link social, economic and demographic understanding to the pattern of health resource utilization. The health resource demand models, part of the overall project, describe actual and expected (normative) utilization of health resources. The differences between the two can then be analyzed by the health planner working with the knowledge supplied by the social area analysis.

The project also sets the upper and lower bounds for this type of health planning analysis. Extensive analyses can also be made at the regional (or major subregional) scale. At the lower end of the spectrum, the analysis can be done with a particular community and its utilization pattern. When the limits of social area analysis are reached, health planners can use methods such as detailed survey work to interpret and understand problems in the utilization of health care resources.

PART II

THE USERS MANUAL FOR HEALTH PLANNERS

II. THE USERS MANUAL FOR HEALTH PLANNERS

The preparation of small area population projections requires significant staff, data, and computer resources. Careful attention must be given in the initial design steps before data and computer programs are collected. The steps are: (1) defining the region, (2) developing regional population projections, (3) defining the social areas, (4) allocating the regional population projections to the social areas.

1. THE REGIONAL APPROACH

The study region should be a single, functional unit carefully selected to include all of the daily economic and social activities of the population within it. From the health resources view there is always overlap in the market or service areas of the health provider institutions. To carefully study the relationships between supply and demand, the region must be large enough to encompass the interaction. From the demographic viewpoint there are three reasons to select a larger region. First, unless all social and economic activities are covered, the socially defined subareas will be incorrect. As the study area becomes a smaller part of the actual region, the errors in the definition of proper subareas increase. The introduction of error is particularly high when an urban area is fragmented by ignoring some suburbs or secondary cities. Secondly, migration is difficult to study at a small area level. With a larger, social and economically defined region the relationships between economic trends and migration can be examined for clues to the future. Lastly, much demographic information is available only at the larger scale. Migration and economic data is generally not available for small, sub-county areas. Births and deaths are normally reported at the county level.

The simplest criteria to bound the region is the daily commuting field. Using census journey-to-work information aggregated to counties the Bureau of Economic Analysis (BEA) of the Commerce Department has delineated a set of 173 regions which are their reporting basis for economic information. These areas are the ideal units but are very large. The most convenient unit that is smaller, but still meets the above criteria, is the Standard Metropolitan Statistical Area (SMSA). The full description of the SMSA's and aggregate statistics are in the County and City Data Book published every five years by the Census Bureau. The SMSA is generally an aggregate of counties that act as a single economic unit. The planner should select as the minimum region the SMSA that contains the central city. Any other SMSAs that border the central city SMSA should also be included to minimize the cross-commuting activity. Other criteria, such as density or political boundaries, may constrain the inclusion, however. For the example used throughout the manual, the San Francisco Bay Area includes the following four SMSAs: San Francisco-Oakland, San Jose, Santa Rosa, and Vallejo-Napa.

2. DEVELOPING REGIONAL POPULATION PROJECTIONS

As described in the first section, the cohort-survival method is selected to produce population projections. Although many methods exist, it is the only one that produces a detailed age/sex breakdown needed for health planning. The cohort-survival method is a proven technique with reasonable data requirements. The Bureau of the Census recommends it and provides training sessions in the use of the method. In its simplest form the method can be performed by hand or programmed to run on a small computer. Simple enhancements to the model expand its usefulness from producing only age/sex projections to include projections of labor force size and composition, number and size of households, and school and group quarters population. This latter, more comprehensive version is described later in detail.

The availability of data and personnel will influence the choice of options in the cohort survival model. In its simplest form the model is a straightforward population accounting procedure tallying births, deaths, and demography. The necessary demographic principles and calculations are described below. There is a considerable range in data requirements as the depth of analysis and various options are selected. The resources necessary to collect and prepare data should influence the model options selected as early as possible in the projections task so that none are wasted. An option requiring large volumes of data will consume a great deal of resources with results that may, or may not, justify the effort.

The first results of the projection model should be examined closely. As described in a later section, a projections review committee was established, made up of state, regional and local planners. Based on their comments changes were made and the model rerun. Their involvement helped in the data collection phase also. During the review process all participants became aware of the model's strengths and weaknesses. The review process is an effective teacher for all concerned.

Once published the projections must not be allowed to rest. Regular re-evaluation of the population projections should be carried out. Many different characteristics can be considered "barometers" of population change. Some of these, local birth records, school enrollments, building permits, driver's license address changes, are available annually and can be used to check the performance of the model in the short run.

In the following sections, the results of ABAG's efforts in each step of the regional and small area population projection process are discussed. From this introduction the reader should begin to accumulate requirements needed to implement such a system. The details of ABAG's experience provide a base to be expanded upon.

An Overview of the Regional Population Projection Model

The ABAG Population Projector and Labor Estimator (APPLE) is a system of computer programs. Five independent programs transform demographic parameters into data inputs for the main program, a cohort-survival model.

Modular Design

Five data generators deal with changes in specific demographic parameters. Program FERT examines beginning year and assumptions about projected end-year fertility levels and birth timing, to compute cohort-specific rates for intermediate years. Beginning and end year mortality rates are processed into needed intermediate year survival rates by the program SURV. Similarly, the programs LFPR and HHEAD process beginning and end year labor force participation rates and household headship rates into intermediate year rates. The outputs of KGEN involve other demographic parameters such as size of the military population, school participation rates and group quarters rates, which in the current design are held as constant. KGEN generates the constant assumption data file as well as the base year data file. The five generators produce written reports and input files for the cohort-survival model, CORE. Complete descriptions for each of the six programs, input, techniques, are in later sections.

The modular design allows the analyst to examine the implications of a shift in a specific demographic parameter quite easily. Unlike previous regional demographic models, large card decks are not required. No longer does one parameter change involve interleaving many replacement cards. Instead, only a few generator input cards need be changed quickly, and the cohort-survival model rerun.

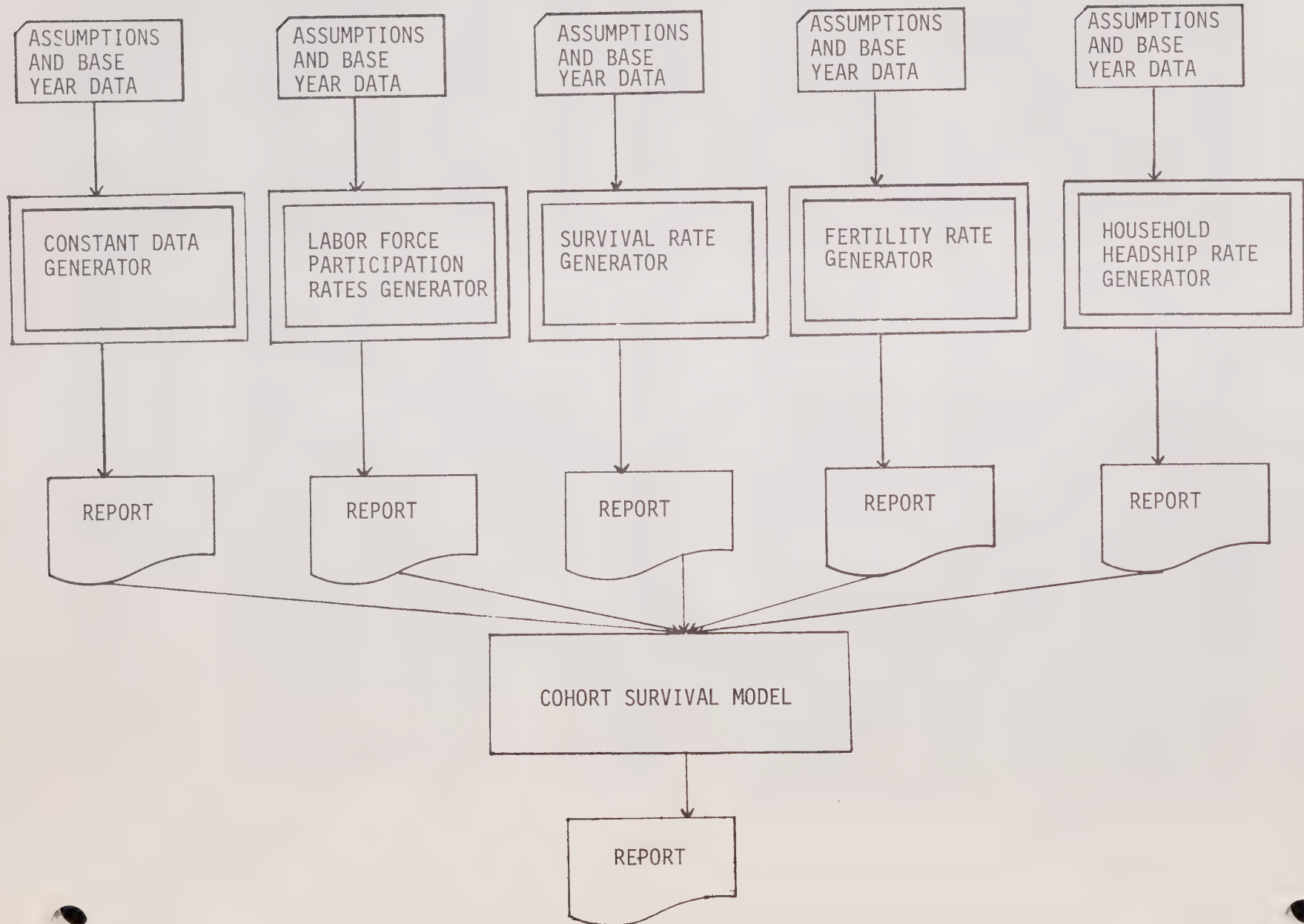
Uses of APPLE

The APPLE system produces, for alternative assumed conditions and policies, demographic and economic projections at five-year intervals. The specific outputs of the model include:

1. Total household and group quarters population by age and sex;
2. Labor force by age and sex;
3. School enrollment for five grade levels, by age and sex;
4. Households by age and sex of head;

Since the future values of the factors producing population and employment changes are the result of analysis of current data (i.e., the model inputs), it is often desirable to produce a range of projections reflecting a reasonable range of the inputs. As a result, a primary objective in developing the model was to produce a technique that would easily generate alternative projections in response to varying assumptions.

FIGURE 2



CORE: The Cohort Survival Model

Purpose

CORE, the demographic submodel, produces, for each age and sex group, conditional projections at five-year intervals of total population, household population, group quarters population (for three categories), households (by age and sex of head), labor force, and school enrollment (for five grade levels). In addition, numerical and percent population changes by age-sex groups are calculated, as well as the population age-sex pyramid, if desired. The model utilizes the cohort-survival technique. Certain population subgroups having unique characteristics (military in-service and their dependents) are read into the model separately, and are involved only in the totaling calculations.

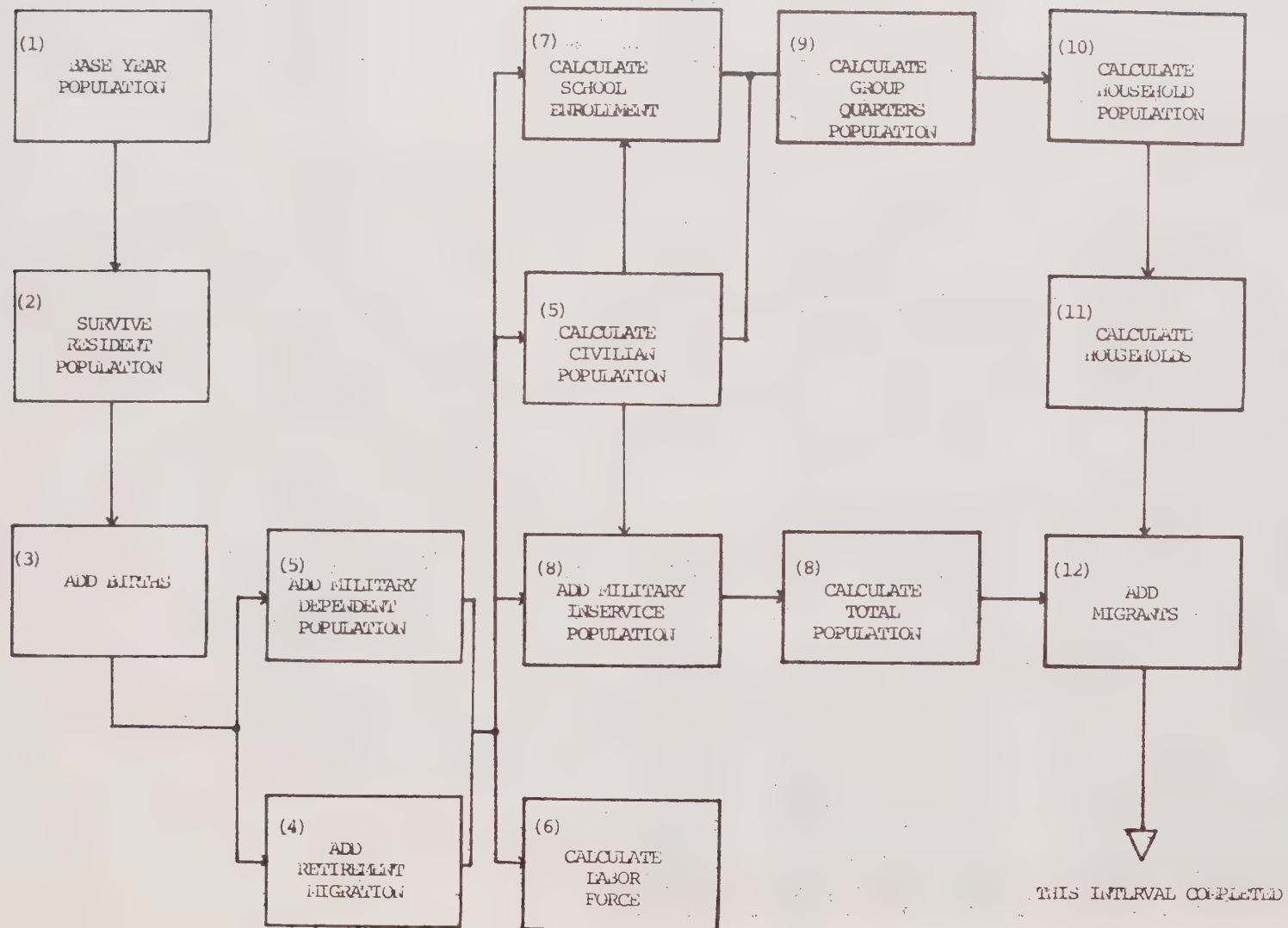
CORE is designed to produce medium- to long-term demographic conditional projections for an economic region. The modular structure of the system allows for adjustments to be made to the early projection cycles. It is recommended, however, that the model be run for projection periods of 20 years or more to fully appreciate the affects of the demographic assumptions. The model works best in regions where the number of workers commuting into or out of the region is insignificant. Such regions include many Standard Metropolitan Statistical Areas (SMSA), all Bureau of Economic Analysis areas (BEA), and in some cases, counties.

Certain data must be prepared for each projection interval. These include forecasts of military in-service and dependent populations, as well as net migration flows. In addition, a number of age-sex-specific ratios are projected by the generators as input to CORE:

1. Fertility rates by FERT
2. Survival rates by SURV
3. Labor force participation rates by LFPR
4. Household headship rates by HHEAD
5. Constant data by KGEN, includes school participation rates (for nursery, kindergarten, elementary, high school and college) and group quarters rates (for military barracks, college dormitories, and other group quarters).

FIGURE 3
CORE FLOW DIAGRAM

(The circled numbers refer to the steps in the text)



Procedure

The following first provides an explanation of variable names followed by a detailed discussion of the model structure and its system of equations. Refer to Figure 3 for the description of model flow.

Explanation of Variable Names

1. Subscripts, Superscripts, and Delta Conventions

- a. For demographic variables, subscripts represented by "i" refer to age groups. To simplify the presentation in this section, reference to sex (male-female) has been omitted; variables with the "i" subscript should be interpreted as applying to both males and females, unless specifically identified as applying to only one sex group. For most variables the subscript i will take values 1 to 18, representing the following 18 five-year age groups: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+. A few variables use fewer age groups; these are specifically noted in the text.
- b. Superscripts refer to time: base year variables have a superscript of "o", projection year variables have a superscript of "t". Certain variables, such as survival rates, apply to a time interval, rather than a point in time; such variables have the superscript "o,t". Since the model always operates with 5-year projection intervals, "o,t" always refers to a 5-year period.
- c. Variables representing a change from one time period to another are designated by " Δ " preceeding the variable's name. For example, Δ POP represents the change in population from base year, o, to the projection year, t.

2. Variable Names

The variable names used in the section, in alphabetical order, are:

- ASFR: Age-specific fertility rate; i.e., the annual number of births per 1,000 women in each childbearing age group.
- AVPOP: Average (female) population in an age group during a projection interval.

BIR:	Births occurring during a projection interval.
CIVPOP:	Civilian population.
ENROL:	Total school enrollment.
GQC:	Population living in college dormitories.
GQM:	Population living in military group quarters (barracks).
GQO:	Population living in other group quarters.
GQRC:	Fraction of college students living in dormitories.
GQRM:	Fraction of military in-service population living in barracks.
GQRO:	Fraction of civilian population living in other group quarters.
HH:	Households.
HPOP:	Household population.
HR:	Household headship rate; i.e., fraction of population in age group that is the head of household.
HS:	Household-serving employment.
HU:	Total housing units.
JPE:	Average number of jobs held by each worker.
LF:	Civilian labor force.
LFA:	Total civilian labor force (available in the demographic sector).
LFPR:	Civilian labor force participation rate; i.e., the fraction of civilians in any age group who are employed or actively seeking employment.
MIL:	Military in-service population.
MILDEP:	Military dependent population.
MP:	Fraction of total employment-related migrants in age group.
PCTPS:	Fraction of total school enrollment in public schools.
PCTM:	Fraction of all births which are male.
PS:	Public school employment.
PSC:	Fraction of college enrollment attending state institutions.

RESPOP:	Resident population, i.e., total population excluding military in-service and dependent population.
RETMIG:	Retirement-related migrants.
SC:	State higher education employment.
SEC:	College enrollment
SEE:	Elementary school enrollment.
SEH:	High school enrollment.
SEK:	Kindergarten enrollment.
SEN:	Nursery school enrollment.
SPRC:	College participation rate; i.e., fraction of population in age group attending colleges and universities (excluding part-time students).
SPRE:	Elementary school participation rate; i.e., fraction of population in age group enrolled in elementary school and kindergarten.
SPRH:	High school participation rate; i.e., fraction of population in age group enrolled in high school.
SR:	Age-specific survival rate; i.e., the probability of an individual surviving for 5 years.
TB:	Total births occurring during a projection interval.
TBF:	Total female births.
TBM:	Total male births.
TER:	Total employment
TMIG:	Total migrants.
TPOP:	Total population.
TSEC:	Total enrollment in state colleges and universities.
U:	Unemployment rate.
VR:	Vacancy rate; fraction of all housing units which are unoccupied.
Y:	Time index; the difference between the base year and projection year.

Steps In Each Forecast Interval (See Figure 3)

1. Base Year Population

Total base year population (including military or any other "special" population) is entered by age and sex categories. There are 18 five-year age groups beginning with ages 0-4 and running through 85 and over.

Next, total base year military in-service and dependent populations are entered by age and sex. These are combined into a single group consisting of the sum of in-service military and military dependents.

Base year resident population is then calculated by subtracting military population from total population:

$$\text{RESPOP}_i^0 = \text{TPOP}_i^0 - (\text{MIL}_i^0 + \text{MILDEP}_i^0) \quad (1)$$

2. Survive Resident Population

The base year resident population expected to survive to year t is then calculated by multiplying resident population by the survival rate:

$$\text{RESPOP}_i^{t'} = \text{RESPOP}_{i-1}^0 * \text{SR}_{i-1}^{0,t} \quad (2)$$

Notice that this process involves moving the RESPOP to a higher age group, from " $i-1$ " to " i ".

The superscript, t' , indicates that the value of RESPOP at this point is not final, and will be modified in subsequent steps.

3. Add Births

Total births are determined by multiplying the age-specific fertility rate by the average number of women in age group i during the (five-year) projection interval. The average number of women in age group i is the mean of the women in age group i in the base year and the forecast year, or algebraically:

$$\begin{aligned} \text{AVPOP}_i^{0,t} &= \left[\frac{1}{2}(\text{RESPOP}_{i-1}^0 + \text{RESPOP}_i^{t'}) + \frac{1}{2}(\text{RESPOP}_i^0 + \text{RESPOP}_{i+1}^{t'}) \right] \frac{1}{2}, \text{ or} \\ \text{AVPOP}_i^{0,t} &= \frac{1}{4} [\text{RESPOP}_{i-1}^0 + \text{RESPOP}_i^{t'} + \text{RESPOP}_i^0 + \text{RESPOP}_{i+1}^{t'}], \end{aligned} \quad (3)$$

where AVPOP is the average number of women in age group i during the projection interval 0 to t ; RESPOP^0 and RESPOP^t are base year and projection year populations as defined in equation 1 and 2, above.

Births for women in each childbearing age group i are then calculated by applying the age-specific fertility rate:

$$\text{BIR}_i^{0,t} = 5(\text{ASFR}_i^{0,t}/1000) * \text{AVPOP}_i^{0,t} \quad (4)$$

Total births are calculated by summing births over the 6 childbearing age groups:

$$TB^{0,t} = \sum_{i=4}^9 BIR_i^{0,t} \quad (5)$$

Total births are then distributed by sex:

$$TBM^{0,t} = PCTM * TM^{0,t} \quad (6)$$

$$TBF^{0,t} = TB^{0,t} - TBM^{0,t} \quad (7)$$

Finally, to account for infant mortality, births are multiplied by the infant survival rate to derive the resident population for the first age group (age 0-4):

$$RESPOP_i^t = TB^{0,t} * SR_{bir}^{0,t} \quad (8)$$

5. Add Military Dependent Population and Calculate Civilian Population

Civilian population is calculated by adding military dependent population (an input) to resident population calculated in (8), above:

$$CIVPOP_i^t = RESPOP_i^t + MILDEP_i^t \quad (9)$$

6. Calculate Labor Force

Civilian labor force is determined by multiplying civilian population by labor force participation rates:

$$LF_i^t = CIVPOP_i^t * LFPR_i^t \quad (10)$$

7. Calculate School Enrollment

School enrollment is calculated for five grade levels: (1) nursery, (2) kindergarten, (3) elementary, (4) high school, and (5) college. School enrollment for each grade level is calculated by multiplying the age-specific school participation rate for that grade level by the civilian population:

$$SEN_i^t = CIVPOP_i^t * SPRN_i^t \quad (11)$$

$$SEK_i^t = CIVPOP_i^t * SPRK_i^t \quad (12)$$

$$SEE_i^t = CIVPOP_i^t * SPRE_i^t \quad (13)$$

$$SEH_i^t = CIVPOP_i^t * SPRH_i^t \quad (14)$$

$$SEC_i^t = CIVPOP_i^t * SPRC_i^t \quad (15)$$

Only the first six age groups are assumed to have school enrollment, that is $i = 1, 6$.

8. Add Military In-Service Population and Calculate Total Population

Total population for the first iteration is calculated by adding military in-service population (determined exogenously and supplied as an input to the model) to civilian population:

$$TPOP_i^t = CIVPOP_i^t + MIL_i^t \quad (16)$$

9. Calculate Group Quarters Population

Group quarters population for the following three groups is calculated: (1) military barracks, (2) college dormitories, and (3) other. Group quarters in military barracks is calculated by multiplying the military population by military group quarters rate:

$$GQM_i^t = MIL_i^t * GQRM_i^t \quad (17)$$

Group quarters population in college dormitories is determined by multiplying college enrollment by the college group quarters rate:

$$GQC_i^t = SEC_i^t * GQRC_i^t \quad (18)$$

Other group quarters population is calculated by multiplying civilian population by the other group quarters rate:

$$GQO_i^t = CIVPOP_i^t * GQRO_i^t \quad (19)$$

For all of the above group quarters calculations, only the following 10 age groups are used: 0-14, 15-19, 20-24, 25-29, 30-34, 35-44, 45-54, 55-64, 65-74, 75+.

10. Calculate Household Population

Household population (for each of the 10 age groups listed above) is calculated by subtracting group quarters population from total population:

$$HPOP_i^t = TPOP_i^t - (GQM_i^t + GQC_i^t + GQO_i^t) \quad (20)$$

11. Calculate Households

Households, by age of head (using 10 age groups), are calculated by multiplying household population by household headship rates:

$$HH_i^t = HPOP_i^t \cdot HR_i^t \quad (21)$$

12. Add Migrants

Total migrants, as input to CORE, are allocated according to the distribution of migrants. As a final computational step, these migrants are added to the cumulative total population:

$$TPOP_i^t = TPOP_i^t + MP_i^t \quad (22)$$

The demographic data computed by the generators are described first followed by a discussion of other parameters. The details on preparing model and generator inputs, and on running the necessary computer programs, are in the software documentation section. This section concentrates on the demographic principles and needed data.

Fertility

The study of fertility rates and timing patterns has long been the central concern of demographers. In the following short summary the necessary components for the analysis of an area's fertility patterns are presented. Following the definition of the basic relationships in the study of fertility, an example will be discussed. The operation of the APPLE fertility rate generator, FERT, is explained in the software documentation section.

Demographic Summary

Three indicators form the basis for the computation of age specific fertility rates: the completed fertility rate, births by age of mother, and the timing pattern. Of these three, the completed fertility rate is most often used as the singular descriptive statistic of fertility patterns, and deserves careful discussion.

Over the 30 or so years of childbearing activity, a woman's total birth performance is measured by the completed fertility rate. This is distinguished from the performance of the entire childbearing population for a single year. The completed fertility rate, then, measures the fertility habits of one cohort of women as it "ages" through the system. A period fertility rate, on the other hand, combines a single year picture of each cohort's habits to form an indicator of the current birth rate. The period fertility rate is subject to fluctuation, based on variation in short-term social and economic trends. Less and slower variation is attributed to the completed fertility rate as it deals with generations of women and their whole childbearing cycle.

Births by age of mother for a given year form the basis for computation of base year age specific fertility rates. These births by cohort are divided by the cohort population to obtain the age specific birth rates (ASBR).

Table 1 supports the above explanation of completed and period fertility rates. The data are the Series II level E (2.1 births per woman) from the Bureau of the Census. Note that the computation of period fertility rates is "vertical"; for the completed fertility, diagonal.

The timing pattern is a distribution of births by age of mother. This distribution indicates, for example, what percentage of a woman's total lifetime births will occur when she is between 15 and 19. A simple representative statistic of the overall timing pattern is the mean age of childbearing. A later timing pattern will generally be represented by a higher mean age of childbearing. Computationally, the mean age of childbearing can be used as the lever to adjust the timing pattern.

Table 2, using San Francisco Bay Area data, illustrates the interrelationship of these three statistics.

TABLE 1
CENSUS BUREAU
PROJECTIONS AND ESTIMATES OF
COMPLETED AND PERIOD
FERTILITY RATES

Age Specific Fertility Rates by Year and Age							
Age	Year						
	1970	1975	1980	1985	1990	1995	2000
15-19	65	55	54	54	54	54	54
20-24	163	158	154	154	154	154	154
25-29	131	129	127	126	126	126	126
30-34	72	69	66	64	61	61	61
35-39	34	32	29	26	24	22	22
40-44	9	9	8	7	6	5	5
							2.115
Period Fertility Rate	2.370	2.260	2.190	2.155	2.125	2.110	2.110

Note: To compute the rates sum one age specific fertility rate for each age group and multiply the sum by 5 and divide by 1000. For period rates, sum vertically; for completed rates, sum along the diagonal.

Source: Current Population Reports, Series P-25, No. 470, p. 52.

TABLE 2
SAMPLE CALCULATION OF
AGE SPECIFIC FERTILITY RATES

Projected fertility rate for the year 2000 = 1.70 births/women		
	I Base Year Births by Age of Mother	II Base Year Female Population
15-19	11268	200,000
20-24	28632	215,180
25-29	23436	185,480
30-34	10075	147,090
35-39	3542	132,970
40-44	892	144,000
45+	48	151,990
TOTAL	77893	1,176,710
	Base Year Age Specific Birth Rates (X1000) (I/II)	Base Year Timing Pattern (Distribution Represented by I)
15-19	56.34	.1446
20-24	133.06	.3676
25-29	126.35	.3009
30-34	68.50	.1293
35-39	26.64	.0455
40-44	6.19	.0115
45+	0.32	.0006
TOTAL	417.40	1.0000

Base Year Period Fertility Rate = $.4174 \times 5 = 2.087$

TABLE 2 (continued)

Age	End Year (2000) Projected Timing Pattern for Mean Age of 27.02
15-19	9.63
20-24	22.76
25-29	35.73
30-34	24.15
35-39	6.53
40-44	1.14
45+	.06
TOTAL	100.00

Projected ASBR for 1519 cohort for the year 2000:
 $= .0963 (1.70 \times .20) = \underline{.03274}$, or 32.74 births per
 1000 women

(The completed fertility rate is multiplied by .20 to
 scale the statistic for the contribution of the average
 of the five year cohort.)

Data Sources

There is a strong need for accurate, current data concerning fertility rates. The quality of data varies widely. Listed below are some of the publications used to develop the base year statistics:

1. U.S. Bureau of Census, "Live Births by Race, Previous Births, and Age of Mother 1970", Table 10, California Life Tables, 21-35.

These tables offered a convenient reference to 1970 Census data. Each county is reported separately by White, Black, and other races.

2. California Department of Finance, "Birth Report B2, Total Births by County of Residence by Age of Mother Cumulative Throughout the State 1974."

This reference is given here to indicate possible sources of current birth data. State departments of health also provide such information. Although a small point, mention should be made of the date of the information. Census data uses April 1 as the base date, where the Department of Finance's Population Research Unit uses a July to June fiscal year for reporting.

3. 1970 Census of the Population, Fourth Count, Tabulation 17, "Regional Summary - Population by Age and Sex."

Census data are widely available in published and machine readable form. This particular tabulation is well documented in the Census publications for a variety of geographical units.

In addition to base year data, information regarding projections of fertility data is necessary. Theories and justifications in the following publications were helpful.

4. U.S. Bureau of Census, Current Population Reports, Series P-20, No. 248, "Birth Expectations and Fertility: June 1973, "U.S. Government Printing Office, Washington, D.C., 1974.

The information was useful in consideration of projected total fertility rates. Very young married women (18 to 21) were interviewed, providing insight to the group that would begin and end childbearing activities in the 1970 to 2000 projection period. These are national figures, and represent opinions of a sample of married women, but nonetheless are useful.

5. Niemeier, Richard, and Savage, Joseph, "New Population Forecasts - Implications for Transportation Planning", AMV Tech News Feature, No. 3, (March, 1975).

This publication provided a good deal of historical completed fertility data, and analyzed the current Census Bureau series I, II, and III. The outcome concluded that even the low series III (1.7 completed fertility rate) may be too high.

6. U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 470, "Projections of the Population of the United States by Age and Sex: 1972 to 2020", U.S. Government Printing Office, Washington, D.C., 1972.

Two timing patterns were suggested by the Census Bureau, an early pattern with a mean age of 25.0, and a late pattern with a mean age of 27.8. The tables provided are useful in choosing and developing a projected timing pattern.

Mortality

Mortality is a negative component in the balance of the vital processes. Understanding of the effects of mortality is necessary to the mastery of the cohort survival method of projecting population. The following explanation will only briefly consider the basis of the study of mortality, the life table. The description of the generator SURV in the software section will act as reinforcement of this explanation. Data sources are indicated to aid in the implementation of the method.

Demographic Summary

A population's mortality level can be described in two ways, the crude death rate, or the mortality rates by age and sex. The crude death rate is simply the total number of deaths that occur during one year among a population per 1000 total population as of the middle of the year.

$$\text{Crude Death Rate} = \frac{1000 * \text{total deaths for year } t}{\left[\frac{\text{population year } t + \text{population year } (t-1)}{2} \right]} \quad (23)$$

At this point, the mid-period timing of the scaling population must be emphasized. To divide by either the beginning period or ending period population would either overstate or understate the population's crude death rate respectively.

The total effect of mortality is masked by the crude death rate statistic, however. The population's age and sex structure must be considered in the understanding of its mortality pattern. For example, two populations with the same crude death rate could have markedly different age and sex structures and, as a result, widely differing population growth patterns. (Consider the different effect of a high infant mortality rate, versus a high incidence of death to women in childbirth, for example.)

Age and sex specific mortality rates offer the sensitivity to the population that is required by the cohort survival method. The computation of mortality rates by age and sex is similar to that of the crude death rate with the addition of an age/sex subscript.

$$\begin{array}{l} \text{Mortality} \\ \text{rate} \end{array} \begin{array}{l} \text{age } i, \text{ sex } j \end{array} = \frac{\text{total deaths for age } i \text{ and sex } j \text{ for time } t}{\text{population for age } i \text{ and sex } j \text{ for time } \frac{(t+(t-1))}{2}} \quad (24)$$

The life table is based on age and sex specific mortality rates. This mathematical model reveals the patterns of mortality among a population for a particular time. By setting up a hypothetical cohort of 100,000 people, and subjecting that group to the accepted probabilities of dying, certain other valuable information is obtained. Indications of longevity are given, as well as a tracing of the lifetime of this hypothetical group.

The life table is best described with an example, by defining its parts and examining how these parts combine to complete a population's mortality picture. Refer to Table 3 for a column-by-column explanation.¹

Column 1: Age Interval ($x, x+n$). The age intervals in Column 1 are the intervals between two birthdays. This column indicates the years of life to which each line on the life table refers.

Column 2: Proportion Dying. This column shows the proportion of the cohort who were alive at the birthday indicating the start of the age interval who will die before reaching the end of that age interval. These figures are similar to, but not exactly the same as, age specific mortality rates. The other columns of the life table are derived from these probabilities; it is the quality of these data that determines the validity of life table results.

Column 3: Number surviving the Interval. If 100,000 births occurred every year, what would be the effect of subjecting this group to the probabilities found in Column 2? Two assumptions are necessarily made: 1) there is no migration, 2) the births are evenly distributed over the calendar year. This is the "stationary" population, so called because the age and sex structure of this population will not change over time.

Column 4: Number Dying in Interval. This column also refers to the stationary population. The probabilities in column 2 are applied to the stationary population by age and sex.

Column 5: Interval Fraction of Life. Of the deaths during the interval recorded in Column 4, there is a contribution made to the total years lived by this group. For example, for the 0-1 age group, the total years lived is equal to those who survived plus .10 of those who died. The underlying assumption is that the average dying person lives about 10% of the one year interval. For the 1-4 group, .39 of the 4-year interval is the average life span of those who died. These fractions are necessary to the computation of the probabilities in Column 2, the number of years lived, and indirectly form the basis of years of all lives remaining, and years of life remaining. Standard demographic tables indicate approximate values for this entry in the life table.

Column 6: Number of Years Lived. As explained for Column 5, these figures indicate the total years lived for the group in the age interval. These total years (note the distinction between years and people) lived for the 0-1 group are the survivors plus the deaths scaled by the average fraction of interval lived. (The .10 rate reveals the timing of mortality for this

¹The following explanation of the life table paraphrases the work of Donald J. Bogue, Principles of Demography. Bogue's work is reformatted to apply to the life table generated by SURV.

TABLE 3

LIFE TABLES FOR BASE YEAR 1970

MALES

AGE	FRACTION DYING IN INTERVAL	NUMBER SURVIVING INTERVAL	NUMBER DYING IN INTERVAL	INTERVAL FRACTION OF LIFE	NUMBER OF YEARS LIVED	YEARS OF ALL LIVES REMAINING	YEARS OF LIFE REMAINING	SURVIVAL RATE
0-1	0.0201807	100000	2018	0.10	98184	6839012	68.39	0.981836
1-4	0.0036329	97982	357	0.39	391057	6740628	68.80	0.995729
5-9	0.0023520	97625	230	0.46	487506	6349772	65.04	0.996454
10-14	0.0022577	97396	220	0.54	486472	5862266	60.19	0.997879
15-19	0.0076597	97176	744	0.57	484278	5375795	55.32	0.995489
20-24	0.0113539	96431	1095	0.49	479364	4891519	50.73	0.989854
25-29	0.0098710	95336	941	0.50	474329	4412155	46.28	0.989496
30-34	0.0101751	94395	960	0.52	469671	3937827	41.72	0.990180
35-39	0.0137623	93435	1286	0.54	464216	3468157	37.12	0.988385
40-44	0.0206961	92149	1907	0.54	456358	3003942	32.60	0.983072
45-49	0.0327733	90242	2958	0.54	444406	2547585	28.23	0.973811
50-54	0.0513691	87284	4484	0.53	425884	2103179	24.10	0.958321
55-59	0.0809338	82800	6701	0.52	397919	1677296	20.26	0.934336
60-64	0.1195563	76099	9098	0.52	358659	1279379	16.81	0.901339
65-69	0.1754422	67001	11755	0.52	306793	920720	13.74	0.855386
70-74	0.2401048	55246	13265	0.51	243732	613928	11.11	0.794450
75-79	0.3421038	41981	14362	0.51	174719	370196	8.82	0.716852
80-84	0.4448694	27619	12287	0.45	104307	195477	7.08	0.596999
85+	1.0000000	15332	15332		91169	91169	5.95	0.466395

group - most deaths occur after only one tenth of one year is lived.) For age intervals of 4 years (1-4) or five years (all the rest) the survivors are multiplied by interval length and added to deaths scaled by percentage of interval lived multiplied by the interval length, i.e.

$$\text{Years Lived} = \text{Interval Length} \times \text{Number Surviving to Next Group} + \text{Deaths} \times \text{Interval Length} \times \text{Fraction of Interval Lived} \quad (25)$$

Column 7: Years of All Lives Remaining. These totals indicate the total years remaining (the product of surviving years and interval length summed for the ages not yet attained). For the 0-1 group this figure is simply the sum of the 19 values of number of years lived found in Column 6. For the 1-4 group the total years remaining are the total years remaining for the 0-1 group minus those years "already lived." In a similar manner, the other 17 age cohort values are computed.

Column 8: Years of Life Remaining. To those surviving the interval (Column 2) must be allocated years of all lives remaining (Column 7). This is done by dividing the years in Column 7 by the survivors in Column 2 for each age group.

Column 9: Survival Rate. For the age groups 5-9 through 80-84, the survival rates are the result of dividing the number of years lived by the 10-14 group, for example, by the number of years lived by 5-9 group. The resulting quotient is the stationary probability of surviving from the 5-9 age group to the 10-14 age group. Please note that the survival rate is not the simple quotient population 10-14 divided by population 5-9. The important difference, and the underlying utility of the life table, is the stabilization of the population to account for fractions of intervals lived. For the age group 0-1, the calculations are performed as follows:

Column 1: age group 0-1

Column 2: Fraction Dying During Interval (QHAT)

$$\text{QHAT} = \frac{\text{Length of Interval} * \text{Mortality Rate}}{(1 + ((1 - \text{Fraction of Interval}) * (\text{Length of Interval} * \text{Rate})))} \quad (26)$$

As is obvious from the above formula, if the fraction of interval lived value were 1.0, QHAT is simply the simple mortality rate adjusted for interval length. Non-unity values for fraction of interval lived result in an appropriate scaling down of the simple mortality rates.

$$\begin{aligned} \text{QHAT} &= \frac{1.02054}{1 + [(1 - .10) 1.02055]} = \frac{.020554}{1.0184986} \\ &= .0201807 \end{aligned} \quad (27)$$

Column 3: For the first cohort this figure is the initial value of 100,000, assumed to be the base population.

Column 4: Number dying in interval is equal to the fraction dying during interval applied to the number beginning the interval (Column 3).

$$\text{NUMBER OF DEATHS} = \text{FRACTION DYING} * \text{NUMBER SURVIVING} \quad (28)$$

$$\text{DEATHS} = .0201807 * 100,000 = 2018 \quad (29)$$

Column 5: Interval fraction of life data are standard demographic estimates and are readily available in publications describing life table development procedures.

Column 6: Number of years lived is calculated by adding survivors to fractions of years survived.

$$\begin{aligned} \text{NUMBER OF YEARS LIVED} &= \text{SURVIVORS} - \text{NUMBER DYING} \\ &+ \text{INTERNAL FRACTION OF LIFE NUMBER DYING} \end{aligned} \quad (30)$$

$$\begin{aligned} \text{YEARS LIVED} &= 100,000 - 2018 + (.10 * 2018) \\ &= 97,982 + 202 \\ &= 98,184 \end{aligned} \quad (31)$$

Column 7: To calculate years of all lives remaining, the number of years lived for each cohort must be calculated. Then the years of all lives remaining for the 0-1 age group is equal to the sum of the number of years lived by all of the 19 cohorts. The value for each of the succeeding cohorts is decreased by the number of years lived by the cohort just previous.

Column 8: Years of life remaining is computed by dividing the years of all lives remaining by the number of people surviving the interval.

$$\text{YEARS OF LIFE REMAINING} = \frac{\text{YEARS OF ALL LIVES REMAINING}}{\text{NUMBER SURVIVING INTERVAL}} \quad (32)$$

$$\text{YEARS OF LIFE REMAINING TO 0-1 GROUP} = \frac{6839012}{100,000} = 68.39 \quad (33)$$

Column 9: The survival rates that are input to the cohort survival model are computed separately for the 0-1, 1-4 groups. These differences occur because the basis for the survival rate computations is the Column 6 data - Number of Years Lived. Below are the two special formulas and the general form:

$$\begin{aligned} \text{SURV 0-1} &= \frac{\text{Years Lived 0-1}}{100,000} = .981836 \end{aligned} \quad (34)$$

$$\text{SURV 1-4} = \frac{\text{Number of Years Lived 1-4}}{4 * (\text{Number of Years Lived 0-1})} = .995729 \quad (35)$$

$$\begin{aligned} \text{SURV 5-9} &= \frac{\text{Number of Years Lived 5-9}}{\text{Number of Years Lived 0-1} + \text{Number of Years Lived 1-4}} = .996454 \end{aligned} \quad (36)$$

$$\text{SURV } i-i+5 = \frac{\text{Number of Years Lived } i \text{ to } i+5}{\text{Number of Years Lived } i-5 \text{ to } i} \quad (37)$$

The life table forms the basis of the SURV data generator. A brief description of the program is in the software documentation section.

Data Sources

To operationalize the SURV data generator, two demographic inputs are required. The base year simple mortality rates (from which mortality probabilities will be calculated), and the projected change in death rates by the end of the projection cycle.

There are life tables available for a variety of geographical bases. California Life Tables are used extensively in this analysis. A series of national mortality data is available from the OASDHI (Old Age Survivors-Disability Health Insurance).

Regional projected change data concerned only crude measures of mortality. National figures were more useful.

U.S. Department of Health, Education, and Welfare. United States Population Projections for OASDHI Cost Estimates, Actuarial Study No. 62, Government Printing Office, December, 1966.

This publication offered three levels of projected change in mortality patterns. These rate changes were chosen for use in the APPLE system.

Migration

Migration flows form the third important component in the study of population growth. Unfortunately, migration data are scarce, and when available, are usually outdated. The APPLE System considers two types of migration: retirement and employment related. Migration is also a possible link between the demographic and economic modeling systems based on a projected employment related migration. The following discussion deals with these two features of migration: as a component of population change, and as a link to other modeling systems.

Demographic Summary

Migration may be studied considering directional flows, i.e., in and out migration, or dealing only with its net effect. Net migration is defined as the signed difference: in-migration minus out-migration. Economists, demographers, and other social scientists have devoted much energy to the modeling of migration's directional flows. For the purposes of our demographic modeling efforts, however, net migration is treated exogenously.

For both employment and retirement-related net migration projection, the age and sex composition of the migrants is an important consideration. This distribution is region-specific. For this reason some effort must be invested to obtain reasonable estimates of the distribution pattern. Useful data sources will be indicated below.

In the absence of specific migration information, the forward survival method can be used to estimate the distribution of migrants by age and sex. This method simulates population change using the cohort survival method. As an example, suppose that only 1960 and 1970 census data are available to study migration. In this unfortunate position, one is forced to assume that all population change in the 10 years between 1960 and 1970 is a result of births, deaths, and net migration. This approach assumes that the number of migrants is constant over the decade, and that future migrants will have the same age/sex composition as in the past. Births to migrants are difficult to analyze since they generally have a fertility lower than the regional level. The 1960 population is "survived" to 1970, using two cycles of the application of representative survival rates. Births are then added, making up two cohorts 0-4 and 5-9. At this step careful study should be made of birth data to estimate this group accurately. The difference between the computed 1970 population and the actual 1970 population is attributed to net migration.

The distribution of migrants by age and sex is each cohort's net migration activity scaled by total 1960-1970 migration. Unfortunately, this past decade has two very different halves, and important effects may be masked by using this span.

Table 4 is an example of this method for employment-related migration.

Estimates of the nine-county region's migration are offered in the Department of Finance publication of "Summary Tables" issued in 1975. Modeling

TABLE 4
 FORWARD SURVIVAL METHOD
 OF APPROXIMATING THE AGE/SEX DISTRIBUTION OF MIGRANTS

1960 Actual Population (Males)		Assumed Survival Rates (60-65)		Births (Male) 1960-1964	
0-4	193,417	0-4	.981836	0-4	197,129
5-9	169,078	5-9	.995729		
10-14	140,515	10-14	.996454		
55-59	80,992	55-59	.934336		
1965 Survived Population & Births		Assumed Survival Rates (65-70)		Births (Male) 1965-1970	
0-4	197,129	0-4	.981836	0-4	181,694
5-9	189,904	5-9	.995729		
10-14	168,356	10-14	.996454		
60-64	75,674	60-64	.901339		
1970 Survived Population & Births		1970 Actual Population		Difference (Actual-Survived)	
0-4	181,694	180,422			-1272
5-9	193,548	211,767			18,219
10-14	189,093	216,221			27,128
65-69	68,208	64,454			-3754

migration trends requires varied data sources. Changes in school enrollments, Medicare statistics, and driver's license address change data are used by the Department of Finance to estimate net migration.

Migration as a Link in the System of Models

Migration is often used as a linking variable when balancing employment requirements with labor force availability. ABAG uses a similar approach. An econometric model projects employment demand by industry. APPLE develops labor force projections based on assumed changes in labor force participation rates. Once the labor force is developed for the projection period, attention turns to the employment (jobs) requirements, which will be projected by the econometric model. Comparison of employment opportunities to available labor force is made, controlling for an assumed long-term unemployment rate. Deficiencies or excess will be corrected by directing the flow of labor force related migrants. This communication between the econometric and demographic models has proved to be a useful means of computing the net migration flow size for different sets of economic and demographic assumptions.

Household Headship Rates

In the following subsections, demographic characteristics will be considered that do not affect the numbers of individuals in the population. School and labor force participation rates, and household headship rates reveal attributes of the total population that are necessary in the presentation of comprehensive projections of population. In this section household formation will be discussed.

Demographics

There are two approaches to the study of residential groupings. The economic approach is the study of the household, or the living unit as the smallest consuming group. Sociological studies emphasize the family as the household unit. The APPLE system incorporates both of these approaches to the study of households.

Bogue defines a household in Principles of Demography as "consisting of all of the persons who occupy a housing unit as a collectivity." A house, an apartment, or even a single room, if occupied as separate living quarters, is considered "a household." Within the household, each individual has status. The status to be considered here is that of household head. Each household can have one and only one head. The choice of head in the family usually defaults to the husband of a family or chief breadwinner. In the case of a household composed of more than one primary individual the choice of head becomes an arbitrary selection. Since the number of heads of households translates exactly to the number of occupied housing units, study of household formation patterns becomes even more important.

As with mortality and fertility there are two levels of indices reported for households. A crude indication is offered by the persons per household statistic. This is the simple result of dividing the non-military, non-group

quarters population by numbers of households. The second and more sensitive level of study is found in the household headship rates by age and sex. Each cohort's number household heads is divided by the non-military cohort population.

Superficial studies of persons per household statistics have indicated that there has been substantial change in household size since 1950. Changes in fertility patterns, a related impact concerning the changing status of women, and economic conditions can be assumed to affect the simple persons per household statistic. Another more complex impact, the changing age structure of the population, also exerts considerable influence on this statistic.

To understand, and, more importantly, to project the numbers of households, attention turns to the household headship rates by age and sex. The role of research on projected patterns of change must be emphasized. Some resource information is offered below but even after studying the information on households at various levels, the amount of uncertainty and the heavy impact of projected change require considerable testing of the assumptions. A useful form for the projected change data is the fractional value: projected household headship rate divided by base year household headship rate by cohort.

Table 5 is an example of the calculation steps involved in developing household headship rates.

Data Sources

Some discussion of data sources should help in the difficult task of projecting change in household formation.

1. U.S. Department of Commerce, U.S. Department of Commerce News, "Census Bureau Reports Decrease in Size of Average Household During Last Five Years", Washington, D.C., February, 1975.

The data and conclusions found in this report support the need for careful study of projected change at the cohort level. The report indicates a drop in the median age of household heads.

2. U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 515, "Estimates of the Number of Households by State, July 1, 1973 and 1972", "U.S. Government Printing Office, Washington, D.C., 1974.

By comparing the state data provided in this publication to the national data found in the previous reference, the relationship of state to national trends may be observed. In California, for instance, household size is below the national average, and the household headship rate for women aged 25-34 is higher than the national rate. This is useful information when developing and calibrating the projected change in household headship rates.

3. U.S. Bureau of the Census, "Regional Summary: Heads of Household 1970 - Source: Sixth County, Tab 1060."

TABLE 5
CALCULATION OF
HOUSEHOLD HEADSHIP RATES

Base Year Household Population (Males)		Base Year Household Heads (Male)	
15-19	193,990	15-19	7,455
20-24	176,074	20-24	90,319
25-29	173,979	25-29	149,606
30-34	144,223	30-34	136,344
35-39	130,907	35-39	125,290
Base Year Household Headship Rate (Males) x100			
	15-19		3.843
	20-24		51.296
	25-29		85.991
	30-34		94.537
	35-39		95.709

This tab details household participation by status for the non-military, non-group quarters population. Both second and fourth counts contain population totals by cohort to act as the denominator in the rate calculations.

4. U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 607, "Projections of the Number of Households and Families: 1975 to 1990", U.S. Government Printing Office, Washington, D.C., 1975.

This publication provided the basis for the projected changes in household headship rates for varying assumed levels of completed fertility. Household headship rates by age and sex are offered. This analysis serves as an excellent starting point in the development of rate change estimates.

Labor Force Participation Rates

Like household headship rates, labor force participation rates do not directly effect population totals. However, within the ABAG modeling system, labor force participation rates provide a valuable indirect link to the econometric modeling system. A well defined and carefully determined labor force is the basis for the determination of the labor force related migrants.

Demographics

Many demographic terms take on various shades of meaning in popular use. For the purpose of this discussion, "labor force" refers to what Bogue calls the "economically active... the workers who are producers of economic goods or services, hence are said to be gainfully employed." Also to be included in this labor force definition are those of marginal work force status. If a worker has worked during the previous calendar year and is seeking employment, or unemployed for seasonal reasons, or seeking employment, this person is considered part of the labor force. Two indicators of the economic participation of a population are treated here. A gross statistic is the unemployment rate. Of those in the labor force, the unemployment rate represents the fraction unemployed and seeking work. This figure is subject to a great deal of seasonal and business cycle fluctuation and does not offer any detail of the economic status of the population. Labor force participation rates offer the required detail. These rates represent the fraction in each cohort that participates in the labor force. Base year data are available from census sources, Bureau of Labor Statistics, and state employment development departments. Having defined the labor force and developed base year rates, it is again time to play the projections game.

The APPLE system allows for cohort by cohort projected changes to the labor force participation rates. Consideration must be given to the current age and sex structure of the population, and any projected changes in fertility patterns and economic conditions. The current age and sex structure of the population may require more active participation of women and retirement aged people to satisfy the demand for workers. This is especially true when each projected completed fertility rate is considered. It is suggested that alternative rates be provided for the various fertility levels assumed.

In the reference cited later, a detail of the 1960 to 1970 changes in labor force characteristics is given. For example, the median age of the workers dropped from 38 in 1960 to about 36 in 1970. (Other assumptions found and supported in this reference will be listed in the Range of Assumptions subsection.)

Table 6 is an example of the calculation of labor force participation rates using Bay Area data for 1970.

Data Sources

It is suggested that some research of recent past and projected changes of economic conditions be performed when developing a set of projected change rates for labor force participation. The first reference was by far the most useful.

1. Johnston, Denis F., U.S. Department of Labor, Bureau of Labor Statistics, "The U.S. Labor Force: Projections to 1990," Monthly Labor Review, July, 1973, pp. 3-13.

This technical report detailed projected change considering such "calculables" as median age of worker, effect of the "baby boom" population of the 1950's, proportion of women, and teenage workers, and participation of aged and male workers. For a detail of the conclusion, refer to the section describing selected APPLE assumptions.

2. U.S. Bureau of the Census, "Census of the Population, 1970 Sixth Count, Tab 1150."

This tab details the labor force status of the resident population by age and sex.

3. U.S. Department of Labor, Manpower Administration, Area Trends In Employment and Unemployment, Washington, D.C., July 1974.

General description of trends is offered, along with useful "labor area" statistics. The data is detailed for 150 labor areas designated by the Department of Labor.

4. Flain, Paul O., U.S. Department of Labor, Bureau of Labor Statistics, "Discouraged Workers and Changes in Unemployment - Hidden Unemployment", Monthly Labor Review, Washington, D.C., March 1973.

This report describes a difficult problem in estimating labor force participation-the marginal worker. It provides good background information and some solutions to the problem of projecting this group's participation.

Other Parameters

The following section describes the following data items: military and military dependents, group quarters rates, school participation rates, and various base year data.

TABLE 6
CALCULATION OF LABOR FORCE PARTICIPATION RATES

Base Year Labor Force (Males)		Base Year Resident* Population (Males)		Base Year Labor Force Participation Rates (Males)	
15-19	76,632	15-19	195,986	15-19	.39101
20-24	131,012	20-24	176,343	20-24	.74294
25-29	153,551	25-29	173,979	25-29	.88351
30-34	135,024	30-34	144,223	30-34	.93622
35-39	125,349	35-39	130,907	35-39	.95754

*By "Resident" population is meant the non-military, non-institutionalized population.

It is in the careful analysis and estimation of these "other parameters" that APPLF becomes a versatile population projection tool. Both military and group quarters population, for example, pose problems to the cohort-survival model. The size and age structure of these groups are determined by influences other than demographic. By including these groups, the projection can represent all members of the population. In the Bay Area each of the group quarters and military related populations represents about 3% of the total population. School participation rates add further utility to the model. While considered stable over time, these rates have widely varying impacts depending on the total population. For each parameter a sample data source reference is provided. In each case, this is the reference used to create the APPIL data set. Also included in this documentation is the method of calculation of values, when necessary. All of these values are input to the generator, KGEN, described in the software section.

Sources and Methods:
Base Year Data/Constant Data

1. Miscellaneous Base Year Data

• Total base year military dependent population

Sources *i*: 1970 Census of the Population, Public Use Sample, Sixth Count, File C, "Total Military and Dependents, Public Use Sample and Sixth Count," for the State of California.

ii: 1970 Census of the Population, Public Use Sample, Sixth Count, File C, "Children by Sex and Spouse of Male Military Head of Household," for the State of California.

iii: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.

Method: To derive the military dependent population totals, appropriate adjustments are made to State military dependents figures as reported in Source *i* above.

The adjustment factor is the fraction a region's non-group quarters military population is of the State's non-group quarters military population (by cohort). The underlying assumption is that military dependent populations are of similar age/sex structure throughout the State. Only non-group quarters military are assumed to have dependents and spouses.

Each cohort's dependent group total for the State is scaled down by the factor described above. (*i* = cohort)

$$\text{factor } X_i = \frac{\text{Region Non-Group Quarters Military}_i}{\text{State Non-Group Quarters Military}_i} \quad (38)$$

$$X_i * \text{State Spouses for Military}_i = \text{Region Spouses for Military}_i \quad (39)$$

$$X_i * \text{State Children for Military}_i = \text{Region Children for Military}_i \quad (40)$$

$$\text{Total Dependents for Region} = \sum_i (\text{Spouses}_i + \text{Children}_i) \quad (41)$$

- Total base year civilian labor force.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.

Method: The total for all cohorts of employed civilians (full time, part time, and those with jobs not at work) and unemployed civilians (except those who did not work during 1970) is computed from the table cited above.

- Total base year school enrollment - Nursery.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: Nursery school figures for each age group for both sexes are added to obtain this total.

- Total base year school enrollment - Kindergarten:

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: The total includes males and females in any age group with kindergarten enrollment.

- Total base year school enrollment - Elementary.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: To derive this total, elementary school figures for each age group for both sexes are summed.

- Total base year enrollment - High School.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: High school totals for both sexes and all age groups are added to produce this total.

- Total base year school enrollment - College

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: College enrollment figures for both sexes and all age groups are included in this total.

- Total base year (occupied) households.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1060, "Regional Summary - Heads of Households 1970," for the nine county region.

Method: To obtain a total of all occupied households in the region, male and female heads of households from all age groups are summed. Heads of households include primary individuals and heads of families (married with spouse present, or other).

- Total base year military group quarters population.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1070, "Regional Summary - Total Population in Group Quarters by Age/Sex/Type of Quarters," for the nine county region.

Method: The military barracks population in 1970 for males of all ages is summed to obtain this total of military group quarters population.

- Total base year college dormitory group quarters population.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1070, "Regional Summary - Total Population in Group Quarters by Age/Sex/Type of Quarters," for the nine county region.

Method: The total college dormitory resident population is computed by totaling male and female dormitory residents for all age groups.

- Total base year other group quarter population.

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1070, "Regional Summary - Total Population in Group Quarters by Age/Sex/Type of Quarters," for the nine county region.

Method: Included in this total of other group quarters population are male and female inmates of institutions (mental hospitals, homes for the aged, or others), residents of rooming or boarding houses, and others.

2. Fraction Male Births

Source: State of California, Department of Public Health: Birth Records for 1970.

Method: For the base year 1970, the total of live births by sex is computed. The fraction of male births is developed by dividing male births by total births.

3. Distribution of Employment Related Migrants

- Sources *i*: 1970 Census of the Population, Second Count, File C, Tab 2, "Regional Summary - Population by Single Year: 1970," for the nine county region.
- ii*: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.
- iii*: Military Dependent Population from item 1 above.
- iv*: "1960 Population (Excluding Military and Military Dependents)," Table 1A, Population Totals File.
- v*: Department of Public Health, State of California, "Abridged Life Tables, California 1959-1961," Bureau of Vital Statistics and Data Processing, September 1963.

Method: To determine the effect of migration, the 1960 population is "survived" to 1970. Any differences in a cohort by cohort subtraction between this derived non-military population (except in the first two cohorts, which must be treated separately) is assumed attributable to employment related migration. Each cohort's difference is divided by the actual 1970 cohort population to obtain the rates. The 0-4 and 5-9 cohort are created by generating births in accordance with the observed 1960-1970 trend.

3. Base Year Population

Source: 1970 Census of the Population, Second Count, File C, Tab 2, "Regional Summary - Population by Single Year: 1970," for the nine county region.

Method: Preparation of these base year figures required the aggregation of the single year totals by sex to the 5 year cohort group totals by sex.

4. Base Year Military Population

- Sources i: 1970 Census of the Population, Public Use Sample, Sixth Count, File C, "Total Military and Dependents," for the State of California.
- ii: 1970 Census of the Population, Public Use Sample, Sixth Count, File C, "Children by Sex and Spouse of Male Military Head of Household," for the State of California.
- iii: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.
- Method: Having developed the military dependent population by age and sex for item 1, (total base year military dependent population), the total military population by age and sex is computed by performing the addition of military in-service and military dependent populations by cohort.

5. Miscellaneous Base Year Data

● Base year occupied households

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1060, "Regional Summary - Heads of Households 1970," for the nine county region.

Method: The total of all occupied households in 1970 is the sum of all household heads.

● Base year school enrollment

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: Total school enrollment is the result of adding nursery, kindergarten, elementary, high school, and college enrollments.

● Base year population

Source: 1970 Census of the Population, Second Count, File C, Tab 2, "Regional Summary - Population by Single Year: 1970," for the nine county region.

Method: The base year population, as described in item 3 (Base Year Population) is summed to produce this total.

6. Military Dependent Population

Sources *i:* 1970 Census of the Population, Public Use Sample, Sixth Count, File C, "Total Military and Dependents," for the State of California.

ii: 1970 Census of the Population, Public Use Sample, Sixth County, File C, "Children by Sex and Spouse of Male Military Head of Household," for the State of California.

iii: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.

Method: (See item 1, Miscellaneous Base Year Data - Total base year military dependents for methodology).

7. School Participation Rates - Nursery

Sources *i:* 1970 Census of the Population, Sixth Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

ii: 1970 Census of the Population, Second Count, File C, Tab 2, "Regional Summary - Population by Single Year: 1970," for the nine county region.

Method: To develop these rates, each cohort's nursery school enrollment is divided by total children in each cohort. Note that both participation totals and cohort population totals include military dependents.

8. School Participation Rates - Kindergarten

Source: (Same as above).

Method: (Same as above, using Kindergarten participants).

9. School Participation Rates - Elementary

Source: (Same as item 7 above).

Method: (Same as item 7 above, using elementary school participants).

10. School Participation Rates - High School

Source: (Same as item 7 above).

Method: (Same as item 7 above, using high school enrollees).

11. School Participation Rates - College

Source: (Same as item 7 above).

Method: (Same as item 7 above, using college participation figures).

12. Fraction in Group Quarters - Military Barracks

Sources *i*: 1970 Census of the Population, Sixth Count, File C, Tab 1070, "Regional Summary - Total Population in Group Quarters by Age/Sex/Type of Quarters," for the nine county region.

ii: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex, and Employment Status," for the nine county region.

Method: Source *i* provides the numerator, i.e., the total military group quarters population by cohort. Total military population data are provided in Source *ii*.

13. Fraction in Group Quarters - College Dormitories

Sources *i*: (Same as Source *i*, item 12 above).

ii: 1970 Census of the Population, Second Count, File C, Tab 1010, "Regional Summary - Year of School in Which Enrolled," for the nine county region.

Method: The college dormitory population for each cohort is divided by the cohort's total college enrollment in 1970.

14. Fraction in Group Quarters - Other

Sources *i*: (Same as Source *i*, item 12 above).

ii: 1970 Census of the Population, Second Count, File C, Tab 2, "Regional Summary - Population by Single Year: 1970," for the nine county region.

Method: Included in this "other" group are the following: residents of rooming or boarding houses, inmates of mental hospitals, homes for the aged, or other institution, and all others. The base population from which the rates are computed is the total population less military and military dependents. The "other" group quarters population is divided by this non-military total population figure for each cohort.

15. Military In-Service Population

Source: 1970 Census of the Population, Sixth Count, File C, Tab 1150, "Regional Summary - Population 14 Years Old and Over by Age, Sex and Employment Status," for the nine county region.

Method: Military In-Service population totals by cohort are found in Source *i*, item 12 above.

Range of Regional Population Projection Assumptions

The initial parameter sets for the APPLE runs are outlined below. The parameters include: fertility, mortality, migration, labor force participation rates, and household headship/size rates. The ranges bracket plausible futures for the Bay Area. They serve here as guides to selecting reasonable ranges and suggest the analyses needed to support the parameters values selected.

Fertility

The fertility assumptions include both birth rates and birth timing patterns.

Birth rates

The strong, unprecedented decline in period fertility at both the national and regional level continues (see figures 4 and 5). The crude birth rate is normally calculated as the number of children born during a year per 1000 total population at the middle of the year. The period fertility rate expresses births in a year in terms of the implied average number of children per woman over her lifetime if, at each year of age, she would experience the current age-specific birth rates. Since the period fertility rate reflects current birth timing as well as the fertility level and age structure of the population, it is the most useful measure of fertility. The crude birth rate can fluctuate about a general fertility time line, and hence is a less useful metric. A population's completed fertility rate represents the childbearing activity of a cohort's total childbearing years.

The total fertility rate for the nation displays a general decline, reaching the previous low of 2.2 in the mid-thirties. The post-war "baby boom" is quite obvious in Figure 4. The recent, sharper fertility decline, both in terms of a crude birth rate and completed fertility, is displayed in Figure 5. Data from 1974 indicates an all time low completed fertility of 1.86 and a 1.44 crude birth rate. The region has also enjoyed a steep decline in births (Figure 6), and has always been ten percent lower than the national rate and lower than state trends. Three possible completed fertility rates are:

High Rate = 2.1

A completed fertility rate of 2.10 represents the replacement level. Both the California State Department of Finance (DOF) and the Census Bureau use this Series E (Series II) level fertility in their projections. It is this completed fertility rate that, when based on 1973 data, yields a projected completed fertility rate for 1974 that is just slightly less than DOF's preliminary estimate for the region. The regional period fertility rate for 1970 was 2.091, computed from Census Bureau data, "Live Births by Race, Previous Births, and Age of Mother, California Counties, 1970." Thus the assumption of a 2.10 completed fertility represents slight growth over the 1970 period level to the year 2000.

FIGURE 4

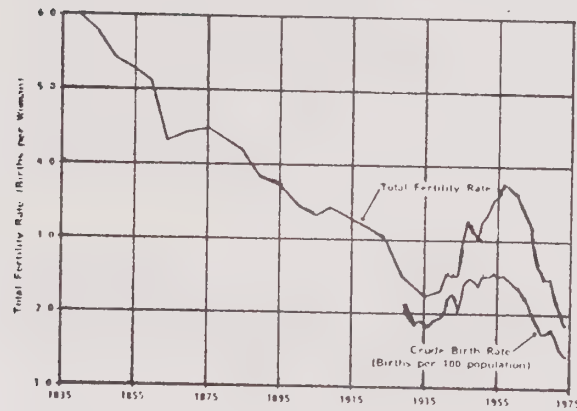
TOTAL FERTILITY RATE,
1835 to 1975

FIGURE 5

TOTAL FERTILITY RATE

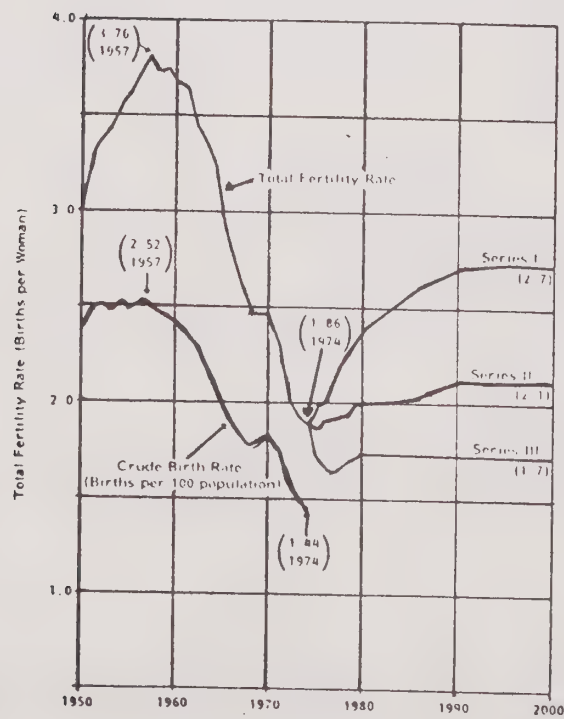
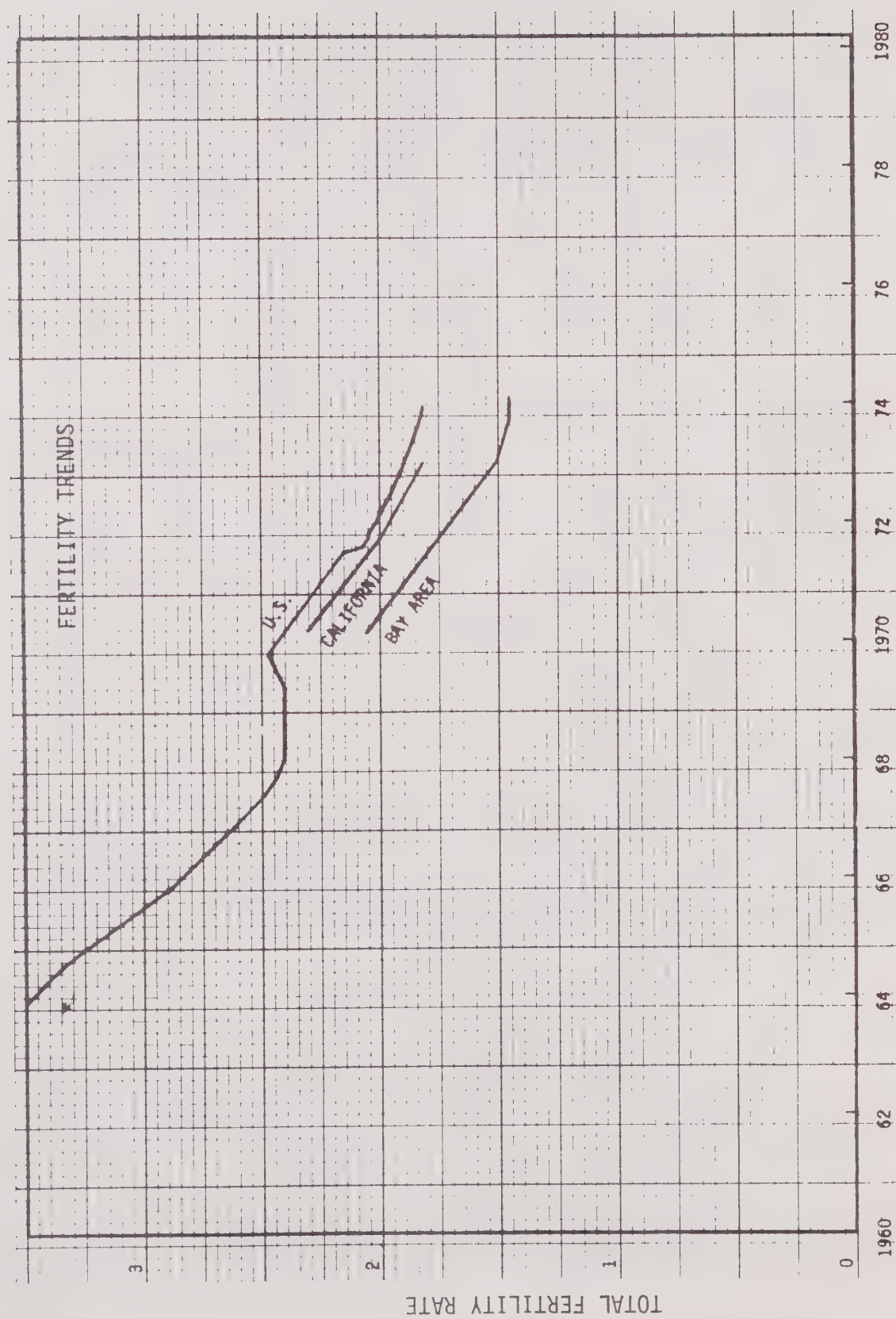


FIGURE 6



In a U. S. Department of Commerce publication "Birth Expectations and Fertility, June, 1973" (P-20, No. 248), the results of a survey of total births expected by married women are reported. Study of the expectations of married women aged 18-24 would indicate the fertility "habits" of the most influential child bearing population for the period 1975 to 2000. For those married women 18 through 21, expectations were up from about 2.23 in 1972 to 2.26 in 1973. Translating expectations to a completed fertility rate results in a slight upward trend to a completed fertility level of about 1.9. Based on this expectation information, an increase to the replacement level seems possible as the high rate.

Medium Rate = 1.7

In the AMV Tech News Feature, "New Population Forecasts-Implications for Transportation Planning", noted earlier, national completed fertility rates are quoted with three Census Bureau predictions for the year 2000. Of those, the authors felt that the 1.7 level (Census Series III) is most realistic. Data for 1974 indicate the nation's period fertility rate has fallen to about 1.86; thus this 1.7 projection would allow for some continuing decline from this period level and then a leveling off, estimated to occur in 1980.

Low Rate = 1.3

A projection of the completed fertility rate at the low level of 1.3 makes possible the study of region's population assuming a continuing decline in the fertility rate. If the current period fertility rate is estimated at between 1.5 and 1.8 (DOF for 1974), this decline to 1.3 by the year 2000 represents a total decline of about 20% in the period fertility rate in the next 25 years. This slow decline serves as marked contrast to the steady, steep trend since 1955, which resulted in a 50% decline in the period fertility in less than 20 years.

Most probably this low fertility rate will be 'powered' by an early timing pattern. This assumption is equivalent to the claim that current timing patterns (mean age 25.10) will remain similar while numbers of births per woman decrease. However, since the timing pattern is dependent on the mean age at childbirth, which is in turn determined by numbers of children, the claim actually represents an inclination on the part of women to postpone the smaller family a little longer.

As yet this low completed fertility rate has not been considered by DOF or the Bureau of Census. In light of the fact that the region's level of period fertility had already dropped below the lowest Series F rate (1.7), this suggested low of 1.3 suggests a reasonable and interesting possibility.

Fertility Timing

The effects of possibly overstated completed fertility rate can be decreased substantially (at least for the early years in the projection period) by propitious choice of a timing pattern.

The Department of Commerce publication "Projections of the Population of the U.S., by Age and Sex: 1972 to 2020" (P-25, No. 470) compares numbers

of births by timing patterns and completed fertility levels. In a ten year period 1970 to 1980, the pair (E, early) actually yields higher results than (D, late). Continuing to the year 2000, such configurations result in a (D, late) total that is only 6% higher than the (E, early).

An early timing pattern would have considerable impact on the '50's baby boom population now in the childbearing years. With continued early timing, a second generation has passed through almost one half of the childbearing years by the year 2000. Late timed fertility, on the other hand, produces an immediate downward trend in total births that climbs as the boom population "ages" to the appropriate level. Such a timing pattern allows for the completion of one generation of activity in a 30 year projection cycle, but finds the next generation just moving into the childbearing years. The comparisons of (E, mid) and (D, late) indicate that the resultant total births differ by only 1%. P-20 series reports offer evidence of an increasing mean age at childbearing. During the past 8 years, more wives responding to the survey felt that their childbearing activities would extend past their 30th birthday. There is a 3% decrease in the reported fraction of expected lifetime births already born by age 29, indicating that more respondents intended to postpone childbearing. Because "more births can take more time" an increase in the completed fertility rate will result in an increase in the mean age at childbearing. This fact suggests that (E, late) would provide reasonable projections.

Two levels of timing are suggested: current, 25.10 (actual mean age at childbearing for the region in 1970) and late, 27.02. This mean age factor will affect total number of births in that they act to create the distribution of births by age of mother, the basis for the period fertility rate generator. The suggested fertility choices are given in Table 7.

Mortality

Mortality rate projection to the year 2000 is accomplished using OASDHI projections, "United States Population Projections for OASDHI Cost Estimates," December 1966. Having established mortality rates for the region (using California Life Tables, 1970), estimates of the proportion change in mortality rates, as presented in the OASDHI publication, were applied to these regional rates. A range of mortality rates is suggested using these OASDHI adjustment figures: the high level assumes no "improvement" and uses the 1970 mortality rates. The medium and low mortality rates are a result of adjustments to regional figures. Medium mortality rates represent one half of the improvement found in the low mortality rates; i.e., if the probability of death for males aged 10-14 were decreased by 40% to represent low mortality rates, a 20% reduction is assumed for the medium mortality rate for this group.

As a measure of the impact of this recommendation, consider the crude death index, the number of deaths per year for all age groups per 1000 population. Using national figures, the range of mortality rates suggested allows for a 0%, 12% and 27% decrease in the crude death rates for males, and 0%, 12%, and 28% for females.

TABLE 7
FERTILITY ASSUMPTIONS

	Completed Fertility Rate	Timing Pattern
High	2.1	late
Medium	1.7	current
Low	1.3	current

TABLE 8
PROPORTION OF BIRTHS BORN BY COHORT

CURRENT*		LATE**
MEAN AGE = 25.10		MEAN AGE = 27.02
15	.19	.19
15-19	14.44	9.44
20-24	36.69	22.76
25-29	30.02	35.73
30-34	12.90	24.15
35-39	4.53	6.53
40-44	1.14	1.14
44+	.06	.06

*Source: census data for region

**Source: calculated

TABLE 9
REGIONAL MORTALITY RATES AND PROJECTIONS

MALES				FEMALES			
Age	Mortality Rate	Percent Change to 2000		Age	Mortality Rate	Percent Change to 2000	
		High	Low			High	Low
0-1	.02018	70.5	41.0	0-1	.01568	70.5	41.0
1-4	.00364	69.7	39.4	1-4	.00276	69.1	38.3
5-9	.00236	70.6	41.3	5-9	.00138	71.1	42.2
10-14	.00226	72.7	45.5	10-14	.00146	72.4	44.9
15-19	.00766	77.2	54.5	15-19	.00350	74.3	48.6
20-24	.01136	77.7	55.5	20-24	.00405	74.8	49.6
25-29	.00962	78.9	57.8	25-29	.00450	75.4	50.9
30-34	.01017	77.7	55.5	30-34	.00585	76.0	52.1
35-39	.01376	76.3	52.7	35-39	.00858	75.3	50.7
40-44	.02070	76.3	52.7	40-44	.01306	74.7	49.4
45-49	.03277	75.9	51.9	45-49	.02059	74.1	48.3
50-54	.05138	75.7	51.5	50-54	.02996	73.5	47.1
55-59	.08099	75.8	51.6	55-59	.04414	74.1	48.2
60-64	.11960	75.6	51.2	60-64	.06072	74.6	49.3
65-69	.17528	76.9	53.8	65-69	.08982	76.0	51.2
70-74	.23999	78.5	57.0	70-74	.13503	76.9	53.6
75-79	.34068	80.0	60.1	75-79	.22158	79.3	58.7
80-84	.45135	81.7	63.5	80-84	.34240	81.1	62.2
85+	1.00000	83.2	66.5	85+	1.00000	82.9	65.9

The crude death rate is influenced by the age and sex composition of the population. The percent change values listed in Table 9 correct for variation in age and sex. Values for the crude death rate may be obtained by weighting the percent change values according to the age and sex composition of the population.

Migration

Examining various sources of net migration estimates reveals a considerable range. Three levels are suggested:

High - 30,000

If the overall national urbanization process shifts again to the major centers, and the economy improves, the migration flows of the 1960's could again develop. The decade's annual average is about +30,000. This corresponds to DOF estimates of the regional share for their E-0 and D-100 projections.

Medium - 15,000

During the past few years, according to DOF, actual net migration magnitude has decreased considerably, in the 14,000-15,000 range. This moderate flow into the region may continue.

Low - Zero

With a continuing decline in net migration zero net migration is the lower bound. This assumption is consistent with increasing growth in middle-size urban centers outside the region, and low growth rates for the major U.S. urban nodes.

Trends

The assumption of one value for net migration is equivalent to decreasing net migration activity. If the total regional population increases, but the net migration flow remains constant, the ratio of the migrants to the total population decreases. For example, the 30,000 annual net migrant flow decreases to about .45% by the year 2000. Such a decreasing trend is consistent with the general relative decrease in net migration flows with increasing total population.

Labor Force Participation Rates

Table 10 presents the age/sex specific labor force participation rates for 1975 through 2000. A different rate schedule is offered for women participants for each completed fertility rate assumed under the fertility discussion above. The projected changes in the U.S. labor force are based on the Bureau of Labor Statistics Report, "The U.S. Labor Force: Projections to 1990" Report 158, 1973.

To 1980:

1. median age decline from 38 in the 60's to 35 by 1980
2. "baby boom" population, currently comprising 20% of the labor force, will grow to 25% by 1980
3. continued increase in the proportion of women workers (from 36.7% in 1970 to 38.5% in 1980)
4. increased participation by teenagers and those aged 20-24
5. increase in the age group 35 to 44 as the large group born between 1945 and 1955 moves into this range

TABLE 10
LABOR FORCE PARTICIPATION RATES*

AGE	Males				Proportion** Change
	1975	1980	1985	1990	1970-1990
15-19	.3889	.3867	.3812	.3805	.9731
20-24	.7394	.7358	.7296	.7233	.9736
25-29	.8812	.8789	.8770	.8752	.9906
30-34	.9333	.9304	.9295	.9276	.9908
35-39	.9536	.9497	.9477	.9467	.9887
40-44	.9459	.9415	.9385	.9366	.9855
45-49	.9308	.9254	.9224	.9205	.9832
50-54	.9251	.9186	.9156	.9126	.9796
55-59	.8676	.8593	.8544	.8515	.9720
60-64	.7217	.7034	.6946	.6896	.9320
65-69	.3406	.3179	.3074	.3004	.8269
70-74	.1762	.1518	.1394	.1314	.6554
75-79	.1132	.0976	.0896	.0845	.6555
80-84	.0680	.0586	.0538	.0507	.6550
85+	.1114	.0960	.0881	.0831	.6554

AGE	Females (Series E, 2.1)				Proportion Change
	1975	1980	1985	1990	1970-1990
15-19	.2955	.3014	.3067	.3113	1.0749
20-24	.6262	.6567	.6723	.6857	1.5111
25-29	.5510	.5902	.5958	.6026	1.1774
30-34	.4955	.5157	.5274	.5338	1.1233
35-39	.5709	.5250	.5381	.5452	1.0976
40-44	.5398	.5508	.5628	.5698	1.0775
45-49	.5483	.5586	.5693	.5762	1.0710
50-54	.5325	.5437	.5543	.5601	1.0742
55-59	.4883	.5014	.5121	.5179	1.0899
60-64	.3843	.3931	.4025	.4077	1.0858
65-69	.1649	.1625	.1625	.1616	.9654
70-74	.0811	.0782	.0752	.0738	.8775
75-79	.0524	.0505	.0486	.0476	.8766
80-84	.0367	.0354	.0340	.0334	.8789
85+	.0515	.0497	.0478	.0469	.8783

*Based on regional data.

**Johnston, Denis F., "The U.S. Labor Force: Projections to 1990,"
Monthly Labor Review, July, 1973, pp. 3-13.

TABLE 10 (Continued)
LABOR FORCE PARTICIPATION RATES

Females (Series F, 1.7)			
AGE	1980	1985	1990
15-19	.3025	.3075	.3066
20-24	.6704	.6891	.7021
25-29	.6064	.6215	.6315
30-34	.5876	.5368	.5464
35-39	.5270	.5412	.5492
40-44	.5515	.5639	.5721
45-49	.5586	.5695	.5765
50+	-as Series E-		

Females (Series H, 1.3) (calculated)***			
AGE	1980	1985	1990
15-19	.3036	.3083	.3020
20-24	.6844	.7063	.7189
25-29	.6230	.6483	.6617
30-34	.5946	.5464	.5593
35-39	.5291	.5443	.5533
40-44	.5522	.5650	.5744
45-49	.5586	.5697	.5768
50+	-as Series E-		

***Calculated under the assumption that a decrease in fertility rates from F to H will have impact on labor force participation rates similar to that caused by the decrease from E to F.

6. "depression cohort" (those born between 1925 and 1934) participation decline
7. decrease in rates for older workers, but when studied in light of rapid increase in this population, this declining rate represents stable total numbers of participants
8. decline in male rates, especially in the 16-24 group where school participation is expected to increase, and among the early retirement group (55+).

To 1990:

1. increase of median age from 35.0 in 1970 to 37 in 1990
2. young workers experience a different mix of male/female distribution
3. slow-down of growth rate of the 25 to 34 group, responding to the declining fertility of the 1960's
4. increase of the 45-54 participation rates
5. declines in 55 to 64 age group as the "depression cohort" population ages
6. slow increase of total participation of the over 65 workers offset by the rapid growth of this population
7. change in the sex distribution of workers: 38.7% women in 1985, 38.8% women in 1990.

Household Headship Rates and Size

Observed Rates:

When examined by age/sex cohort the 1960 and 1970 censuses support some interesting hypotheses. Headship rates for females in all relevant age cohorts increased, with substantial impact observed in the groups aged 20-54. During the decade it seems there occurred some radical change in the composition of households (assuming, of course, that identical methods of determination were used). There are increases in all but one age group of males. Interestingly, this age group 20-24 where the male headship rate declined is where the greatest increase in female headship rates is found.

The Census Bureau reports that the national median age of household heads dropped from 48.1 to 47.3 in the five year period 1969-1974. In the region, the median age of household heads decreased 5.8% from 47.7 in 1966 to 45.10 in 1970. Convolving these facts provides support for the hypothesis that there will be continued growth in these rates to the year 2000.

Examination of data available from the 1950 census provided inconclusive results. Only gross rates of heads of households by sex were available. When compared to similar data for 1960 and 1970, the 1950 rates indicated that substantial change has occurred in the sexual distribution of household heads. For both decades, female headship rates increased, with the slower increase occurring between 1960 and 1970.

Suggested Trends

Simple extrapolation of this decade's growth trends results in patterns that seem too high. Assuming that "some of the same" will continue

through the projection period, some fraction of this pattern may be adopted as estimates of household headship rate change to the year 2000. Study of results obtained using a range of growth "acceptable" indicates that 33 per cent of the rate change observed from 1960 to 1970 provides reasonable figures. Such projected household headship rates may be low when considering decreasing household size and decreasing birth rates. However, further growth past this point yields male categories with one hundred percent rates.

Reports on Household Size

Assume that California rates apply well to the 9-county region. California provides special interest as an area in which to examine the change of character of the household unit. For two reasons California acts as a special case: 1) during the period to be examined (1960 to 1973) California had experienced high rate of increase in the adult population, and 2) in California, there is a higher proportion of primary individuals for all relevant ages. These facts, together with a national trend² that has acted to increase household headship rates within the age groups where these rates are already high, are substantiated in Tables 11 and 12.

Continued research of household headship studies has yielded projections of headship rate changes by varying levels of marital status and household status. A Census Bureau study³ documents numbers of households by type, status of head, and age of head for the estimate year 1974. The three projections of total numbers of households are based on extrapolations to 1990 of changes in marital status and household status observed over the period 1960 to 1974 and on projections of the total population by age and sex.

Since the statistic required for the household headship rate generator is the rate of change to the end of the projection period, the data and projections provided in the study underwent considerable calculation steps. Total change in numbers of households by age and sex of head, was calculated for each cohort. The report documented the increase in total households attributable to "natural growth." For each cohort the natural growth was stripped away, because the headship rate must represent projected change in headship formation only.

Each of the three levels of projected change in household headship rates must be matched with the level of fertility that is consistent with the accompanying projections of change in marital and household status. For example, if fertility were to decline to 1.3, the projected impact on household headship rates would be increased participation of women in particular, and all groups in general, to produce a marked increase in number of households while generating a decline in household size.

²U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 515, "Estimates of the Number of Households by State, July 1, 1973 and 1972", U.S. Government Printing Office, Washington, D.C., 1974.

³U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 607, "Projections of the Number of Households and Families: 1975 to 1990", U.S. Government Printing Office, Washington, D.C., 1975.

TABLE 11
NUMBER OF HOUSEHOLDS

	1973	1972	1970	1960
California	7,175,000	6,972,000	6,573,861	4,981,024
U.S.	68,737,000	67,132,000	63,449,747	53,021,000

PERCENT CHANGE IN NUMBER OF HOUSEHOLDS

	Total 1970-1973	Average Annual Percent Change 1970-1973	1960-1970
California	9.1	2.7	2.8
U.S.	8.3	2.5	1.8

HOUSEHOLDS BY TYPE

	Husband/Wife Households		Non-Husband/Wife Households		
	1973	1972	1973	1972	Percent Change
California	4,543,000	4,476,000	2,632,000	2,496,000	5.45
U.S.	46,651,000	46,061,000	22,086,000	21,071,000	4.82

(Table 11 continued)

¹⁷U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 515, "Estimates of the Number of Households by State, July 1, 1973 and 1972," U.S. Government Printing Office, Washington, D.C., 1974.

TABLE 11 (continued)
DISTRIBUTION OF HOUSEHOLDS

	Husband/Wife			Non-Husband/Wife		
	1973	1972	Percent Change	1973	1972	Percent Change
California	63.32	64.20	-1.39	36.68	35.80	2.46
U.S.	67.87	68.61	-1.09	32.13	31.39	2.36

Source: U.S. Bureau of the Census, Current Population Reports, Series P-25 No. 515, "Estimates of the Number of Households by State, July 1, 1973 and 1972," U.S. Government Printing Office, Washington, D.C., 1974.

TABLE 12

BASE YEAR (1970) HOUSEHOLD

HEADSHIP RATE

AGE	MALES	
	1970	PERCENT CHANGE TO END YEAR
0-5	0	0
5-9	0	0
10-14	.137	0
15-19	3.843	12.70
20-24	51.296	-5.02
25-29	85.991	5.84
30-34	94.537	5.44
35-39	95.709	2.17
40-44	95.709	2.17
45-49	95.509	1.45
50-54	95.509	1.45
55-59	95.328	2.00
60-64	95.328	2.00
65-69	93.737	2.89
70-74	93.737	2.89
75+	86.929	5.88

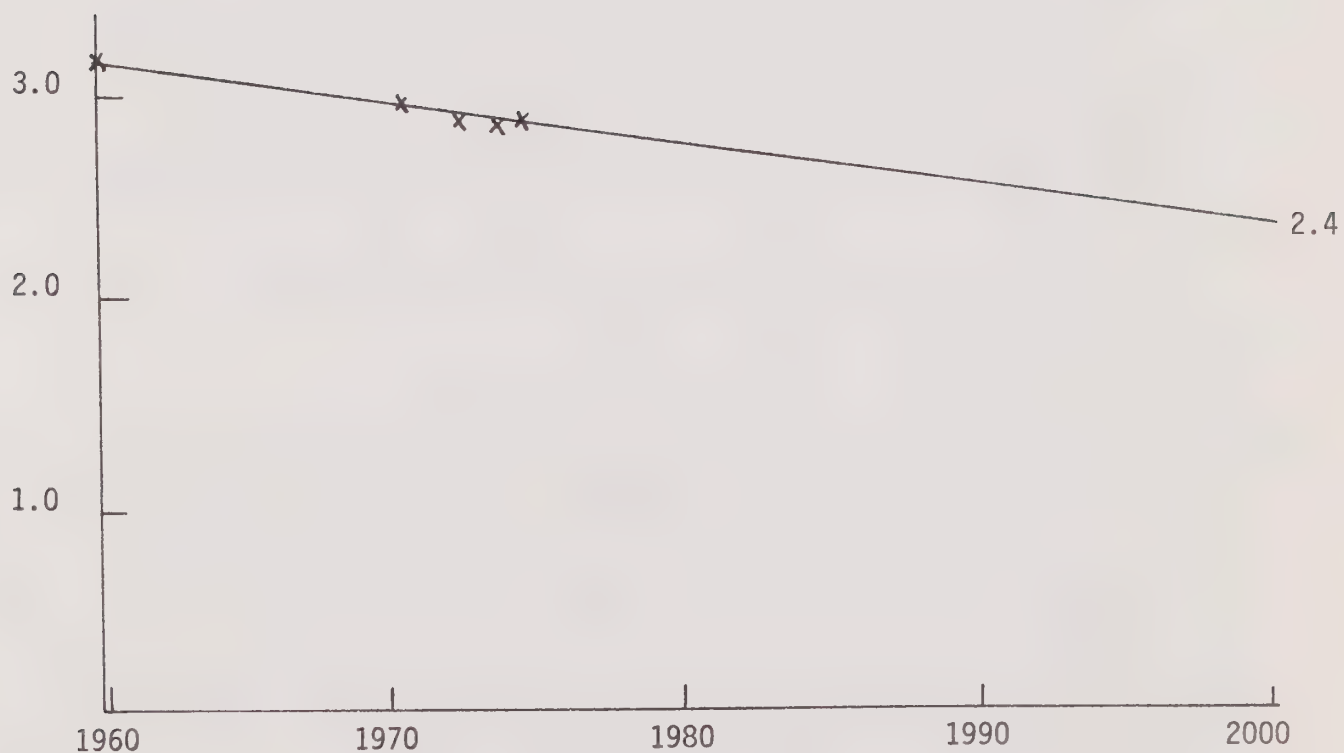
AGE	FEMALES	
	1970	PERCENT CHANGE TO END YEAR
0-5	0	0
5-9	0	0
10-14	.042	0
15-19	2.302	21.16
20-24	15.022	52.51
25-29	16.494	51.18
30-34	16.525	43.70
35-39	19.005	40.26
40-44	19.005	40.26
45-49	25.462	36.82
50-54	25.462	36.82
55-59	32.034	14.86
60-64	32.034	14.86
65-69	42.985	6.27
70-74	42.985	6.27
75+	53.165	12.35

Source: U.S. Bureau of the Census, "Regional Summary Head of Household:
1970 - Source: Sixth County, Tab 1060."

TABLE 13
 PROJECTED CHANGE IN HOUSEHOLD HEADSHIP RATES
 BY THE YEAR 1990

	HIGH	MEDIUM	LOW
MALES	I	II	III
20	1.0204	1.0042	.9241
20-24	1.0107	1.0027	.9776
25-29	1.1099	1.0677	1.0284
30-34	1.1898	1.1227	1.0575
35-44	1.2116	1.1356	1.0623
45-54	1.0351	1.0209	1.0077
55-64	1.0388	1.0229	1.0078
65-74	1.0887	1.0589	1.0290
75+	1.1222	1.0762	1.0323
FEMALES	I	II	III
20	1.0680	1.0261	1.0219
20-24	1.2884	1.1352	1.0930
25-29	1.3682	1.1908	1.1188
30-34	1.3742	1.2102	1.1207
35-44	1.2742	1.1650	1.0885
45-54	.9424	.9787	.9424
55-64	1.0163	1.0118	1.0053
65-74	1.1499	1.0905	1.0484
75+	1.2425	1.1423	1.0782

FIGURE 7
HOUSEHOLD SIZE
Persons/Household
In California



Based on the State of California information

Source: U.S. Bureau of the Census, "Regional Summary Head of Household:
1970 - Source: Sixth Count, Tab 1060."

APPLE Sensitivity Analysis

The many hours of research and analysis that went into the preparation of APPLE assumptions only served to create the need for even further evaluation. Preliminary results of any projections must be studied in relation to both the size and distribution of the resulting populations. Sensitivity analysis is used to observe the effects of various assumptions. The first step is the identification of the critical input parameters of the model or those variables that directly affect the total population outcome. The three parameters, fertility, mortality, and migration were varied over their selected ranges in order to determine the degree of variation in the resulting projections.

Because both household headship and labor force projected change are related to changes in the completed fertility level, a multivariate sensitivity analysis of the interaction of these three variables is most useful. A further challenge to the sensitivity analysis is to determine the affects of migration on household formation and labor force participation. This challenge suggests the need for another multivariate sensitivity analysis. The age and sex structure of the population must be considered in each of these multivariate analyses. The project quickly becomes a monumental task.

Below is described the first level, univariate sensitivity analyses performed on the preliminary APPLE runs.

Sensitivity Analysis Design

To best observe the impact of one variable, the amount of variance external to it must be minimized. The analysis of each of the three parameters involves holding the other two constant. These constant levels were chosen either for their "most probable" nature (fertility and mortality) or for their "ease in computation" quality (migration). Fertility testing allowed the completed fertility rate to assume the values 1.3, 1.7, and 2.1 births per woman. The assumed net migration levels to be tested were 0, 15,000, or 30,000 migrants per year. The mortality rates varied as suggested by OASDHI, from a high level of 40 percent of 1970 mortality to a medium rate of about 65 percent of 1970 mortality to a lowest level of the 1970 rates held constant throughout the projection period. When testing migration or mortality, fertility was defined as 1.7 births per woman. For fertility and mortality migration was set to zero migrants per year. 1970 mortality rates were held constant while testing fertility and migration.

Sensitivity Analysis Findings

The "lever" statistics, births per woman, number of migrants, relative change in mortality rates, complicate the process of comparing resulting totals for the different parameters. For this reason, each of fertility, mortality, and migration is subject to a separate, self-contained analysis.

Fertility

1. Of the three parameters, fertility rate variation produced projections with the middle range size. The range of total population for fertility is 14.02 percent of total, for mortality 5.58 percent, and for migration 22.99 percent.
2. When migration and mortality are held to their constant values, the various levels of completed fertility produce the following simple growth rates from 1970 to 2000: 1.3 births/woman, 2.72 percent; 1.7 births/woman, 7.07 percent; and 2.1 births/woman, 14.02 percent.
3. It is interesting to note the effect of the 2.1 completed fertility rate compared to the 1.7 level. Currently the region's period fertility is 1.55, below the 1.7 completed fertility level. The upturn to 2.1 of the completed fertility level represents a 6.95 percent growth rate difference in total population over the 1.7 level.
4. As with the influence of mortality rate changes, variations in fertility levels result in changes in the distribution of the population by age. Obviously, the 0-9 cohort reflects the level of change in the fertility rate. When the fertility rate is allowed to increase to 2.1 by the year 2000, the "children cohorts" - those groups most affected by changes in fertility rates (0-29 for this projection period) make up 42.5 percent of the total population. At the two other levels of fertility, 1.7 and 1.3, these percentages drop to 37.8 percent and 34.7 percent, respectively, by the year 2000. When comparing these distribution figures to the 1970 base year data, the base year "child cohort", 0-29, is 52.4 percent of the total population. This proves itself to be a poor basis of comparison because the base year distribution is affected by the "baby boom" of the late forties and early fifties. However, the composition of the young population will vary considerably with the increasing completed fertility rate. When assuming the low level of fertility, only about one fourth of the children produced from 1970-2000 are below age nine. At the high level of fertility, this fraction increases to one third. In fact, at the high level of fertility, all three ten year cohorts in the group of children produced share equally in their distribution. The low assumption of fertility produces very different results. In this case the distribution of the 0-9, 10-19, 20-29 mix is 25 percent to 33 percent to 42 percent, indicating a definite decrease in childbearing activity.
5. The effects of the variation in timing pattern must be noted. For the first cycle, the projection representing the lowest fertility level actually produces a higher total population than the projections using either the 1.7 or the 2.1 rate. Here is seen the dual effects of the interpolation algorithm and an early timing pattern. Because end year values are adopted differentially by cohorts to be completed by the year 2000, very little change is demonstrated this early in the projection period (only the 15-19 year olds have adopted end year patterns by this time). Because of the early timing pattern (used

with the low and medium completed fertility assumption), the "baby boom bulge", i.e., the 20-29 year olds, have begun to produce families earlier, and thus effect this cycle's totals. By 1980, however, this timing pattern effect has begun to be swamped by the higher completed fertility rate assumptions.

6. With the assumptions of zero net migration and constant 1970 mortality levels, using the medium and low completed fertility rate, and its accompanying early timing pattern, the population actually experiences decline in the later years of the projection period. When 1.7 is the completed fertility rate, this decline begins in 2000; when 1.3 is the assumed rate by the year 2000, the population begins to decline in 1990.

Mortality

1. When compared to fertility and migration, varying the mortality levels has the least impact on the total population. The range produced represents a fraction .0558 of the total population.
2. The simple base increase of the population is assumed to be the 1.7 completed fertility, zero net migration, and 1970 mortality rates held constant projection of the total population. The impact of declining mortality rates is as follows. If the mortality level were to experience the extreme drop to 40 percent of the 1970 rate, the impact on the total population would be an increase of only 5.32 percent by the year 2000. When isolating the impact of half that decline an increase 2.50 percent by the year 2000 is observed.
3. The conclusion follows that the level of total population is not very sensitive to even radical changes in mortality rates. Some suggestions are offered below to explain this insensitivity.

Although the projected change in mortality varies from optimistic to plausible, both changes occur slowly, attaining the projected level by 1990. Perhaps if the model is run over a longer span of time, the impact of such assumptions would be more apparent.

Both of the projected changes in mortality rates effect the over 60 group slightly more than the rest of the population. This group is small (only 12.7 percent of the population) and is, therefore, relatively powerless in demonstrating the effect of changing levels of mortality.

4. Although the impact of varying levels of mortality on total population is not startling, there are interesting variations in the population distribution by age.

In 1970, approximately 12.7 percent of the population were aged 60 and over. Assuming no change in mortality rates, this group grows to 18.20 percent of the total population by the year 2000. The relative growth in this group is a function of declining fertility (1.7 versus the 2.1 completed fertility rate in 1970) and the fact that in 1970 the 30-59 year olds comprised a large part of the region's population. As this group "ages" to the year 2000, it maintains its position as a relatively large cohort. When mortality levels drop to 40 percent of the 1970

level by the year 2000, the relative growth of the over 60 group is quite dramatic: from 12.7 percent in 1970 to 20.9 percent in 2000. This represents a 14.83 percent increase in the rate of growth of this group when the mortality rates are assumed constant. This fact also indicates that, when considered over a longer time span, changes in mortality levels become more important in the way that both total population and population distribution are affected.

Migration

1. It is not surprising that with a range of 0 to 30,000 net migrants per year varying this parameter generates the widest range of projections, 22.99 percent of the total population in the year 2000.
2. The base assumption of zero net migration, coupled with fertility projected at 1.7 births/woman and mortality held constant at the 1970 level results in a growth rate of 9.79 percent from 1970 to 2000. When each of the alternative assumptions (15,000 and 30,000 per year) is stripped of this simple growth effect, the total change figures are as follows for each of the assumed net migration levels:

0-15,000 per year results in a 11.50 percent increase over base year due solely to the migration.

0-30,000 per year results in a 22.99 percent increase over base year due solely to the migration.

3. Consider the impact of the 1.7 completed fertility rate and the constant mortality level.

If migration is 15,000 per year, then by year 2000 an additional 450,000 people are to be included in the population totals. Actually, 532,300 is the difference between the 2000 projected population with no migration and the 2000 projected population with 15,000 migrants per year. The effect of the 1.7 fertility rate and constant 1970 mortality rates is an excess of 82,300 migrant-related persons. If migration is 30,000 per year, the add-on of expected migrants is 900,000 versus the 1,064,600 observed as the difference between population when there are 0 and 30,000 migrants. An excess of 164,600 is observed as a result of fertility and constant 1970 mortality.

These findings lead to an examination of the migration assumptions as to magnitudes, but more interestingly, as to the distribution of migrants. The employment related migrant population is composed of young people (58 percent of migrants are 29 or younger). This fact sheds some light on the reasons for the high "effective" migrant population. 58 percent of the women who migrate are between the ages of 15-34. More specifically 38 percent of migrating women are within the age groups of 20-24 (17.39 percent) and 25-29 (20.98 percent). The timing pattern adopted for the fertility generator has 25.1 as the mean age of mothers at childbirth. The resulting distribution of births by age reveals that 68 percent of all births occur to women from age 20-29.

As a contrast to the migrant population, the total Bay Area population in 1970 had 52.30 percent of total aged 29 or less. While this is not significantly different from the 58 percent found in the migrant population, the women in the prime childbearing years provide marked contrast. Among the resident population, only 16.9 percent of the women are aged 20-29.

4. This comparison indicates that further study is required to determine the "effective migrant " population. Under the two migration assumptions (30,000 and 15,000 net migrants per year), an excess of 16 percent of the resultant migration population by the year 2000 is composed of offspring of migrants. When considering the subtracting effect of mortality this increase due to fertility is even greater. (However, the effect of mortality on migrants is small in that few employment related migrants are 50 years or older, where mortality has its greatest effect.)
5. Varying the levels of net migration has little effect on the distribution of the population by age for those 29 or under. There is slight impact to the older groups by the year 2000, the greatest variation is found in the 30-49 group with the range of about +2 percent.

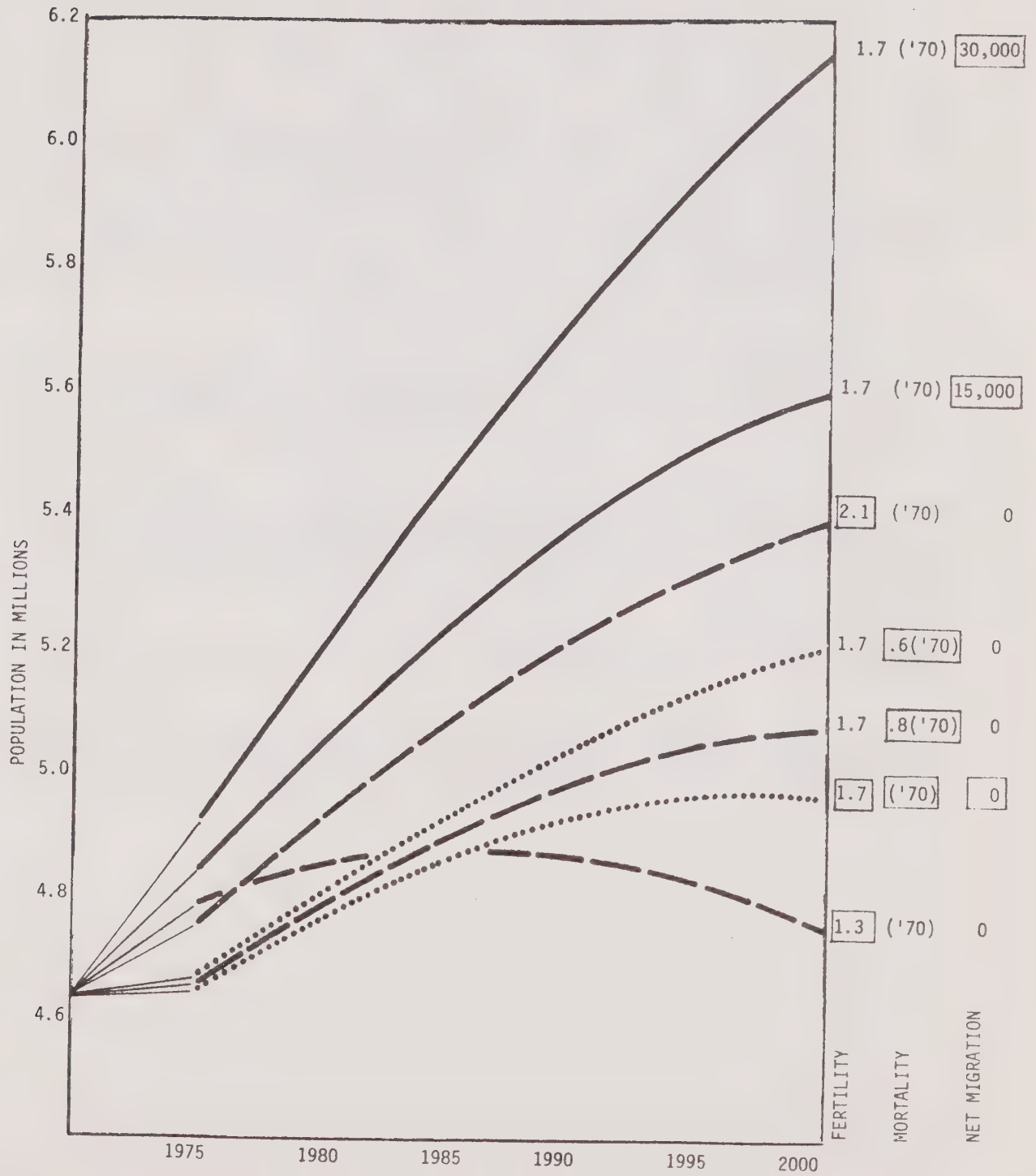
The Role of an Advisory Committee in the Projection Process

During the projection work, the analysts found that an open process was useful. PTAC (Projections Technical Advisory Committee) was established to provide a forum where the development and use of projections could be discussed. The Technical Advisory Committee was composed of planners, demographers, and other technical personnel from the city, county, regional planning departments in the Bay Area and from a few California State Offices. The committee served several purposes. A typical PTAC meeting had a pre-meeting mailout including an agenda and any working papers subject to discussion at the meeting. For example, as research on trends in migration, mortality, and fertility progressed, PTAC members were asked to comment on suggested parameter values. Often the county and city representatives would be able to provide data and insight to support or refute the suggestions. Appreciation of the difficulties involved in the regional modeling effort was a result of one particular session where members were solicited for their suggestions as to how to reconcile APPLE's population and labor force projections with the results of the economic modeling work. All benefited from the open process.

Producing the Final Regional Projections

As a result of the sensitivity analysis, several PTAC meetings and several initial test runs, two demographic assumption sets were selected for the regional model. Each set is called a base case.

FIGURE 8
SENSITIVITY ANALYSIS
(APPLE runs at ABAG)



Each Base Case projection has a different set but not every assumption was changed for Base Case 1 and 2. The major consideration in selecting assumptions was plausibility. Recognizing the uncertainty inherent in projecting the future, different sets of plausible assumptions were selected to project a range of future outcomes. The sets of assumptions for Base Cases 1 and 2 are summarized in Table 14. The major variation among the assumptions is at the regional level. The projections are most sensitive to assumptions about fertility, migration, and economic growth. It should be noted that there is a strong relationship between economic growth and migration. One final consideration in selecting the assumptions was the size of the projection range. It was felt that a wider, rather than narrower, range would be better for planning purposes. However, this criterion was not strictly adhered to for several reasons: 1) there was no general agreement as to what constitutes an acceptable range; 2) the range results from the assumptions, and not vice versa; and 3) it was not possible to maintain a wide range for all of the variables projected.

Using the APPLE model with the two assumption sets results in a Bay Area population of 5.4 to 6.1 million for the year 2000, compared to a 1975 population of 4.8 million. The projected 1990 population range is 5.3 to 5.6 million.

Labor force and employment are estimated to grow at a faster rate than total population. The projected average annual growth rate for labor force through 2000 is 0.9 to 1.3% compared to a population growth rate of 0.5 to 1.0%.

Households are estimated to form a faster annual rate than either population or labor force, 1.4 to 1.6%.

In Base Case 1, total births rise until 1990, due to the assumption of increased birthrates from the 1975 level. Births decline in both Base Cases as the number of women in the high fertility ages (20-29) decreases with the aging of those born during the postwar baby boom.

TABLE 14
SUMMARY OF FINAL PROJECTION ASSUMPTIONS

REGIONAL ASSUMPTIONS	BASE CASE 1	BASE CASE 2
Fertility	1.8 children per woman	1.5 children per woman
Migration	Increasing from 25,000 per year in 1975 to 35,000 per year at 2000	Decreasing from 20,000 per year in 1975 to 10,000 per year at 2000
Mortality	Slightly declining rate	Slightly declining rate
Average Household Size	Declining size consistent with 1.8 fertility level	Declining size consistent with 1.5 fertility level
Economic	Selected industries con- tinue to grow relative to national growth rates	Selected industries con- tinue to grow, but slower than in Base Case 1

FIGURE 9

AVERAGE ANNUAL BIRTHS, DEATHS, NET MIGRATION(TEN YEAR MOVING AVERAGE)

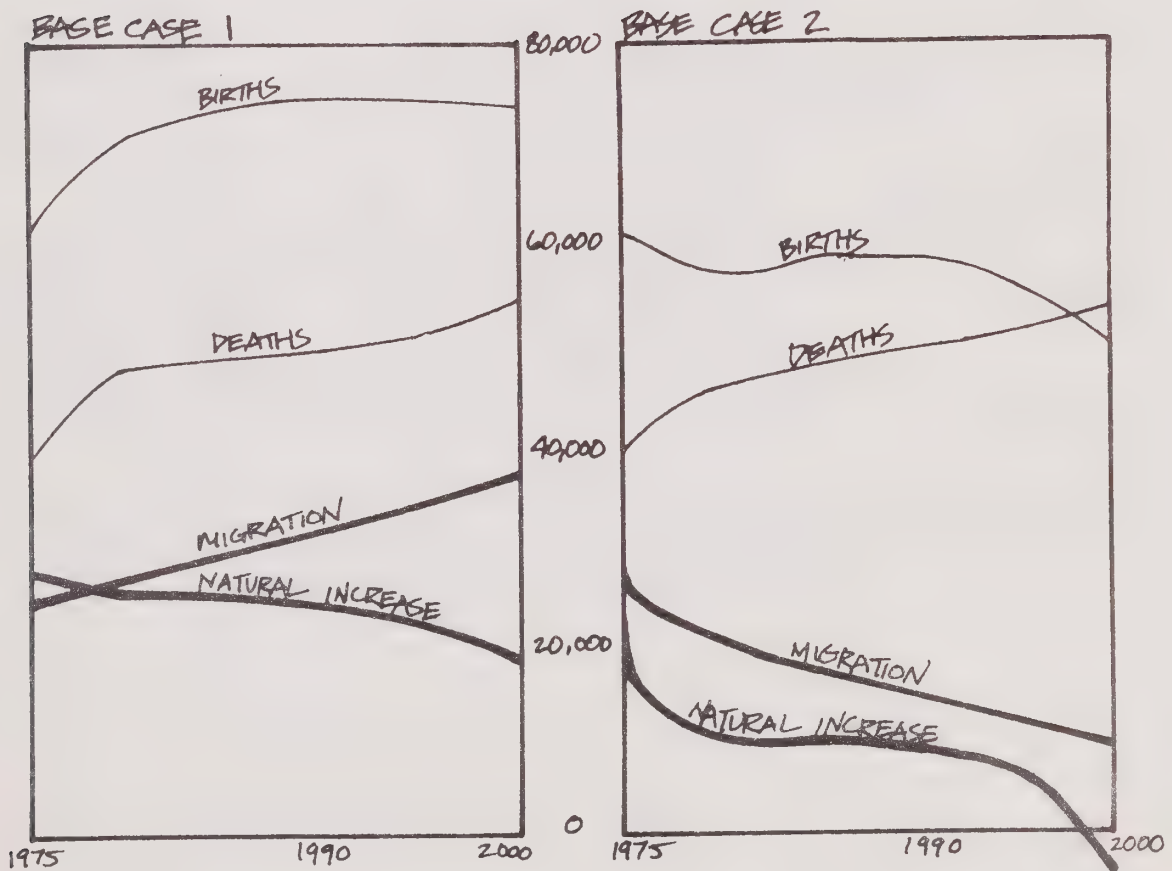
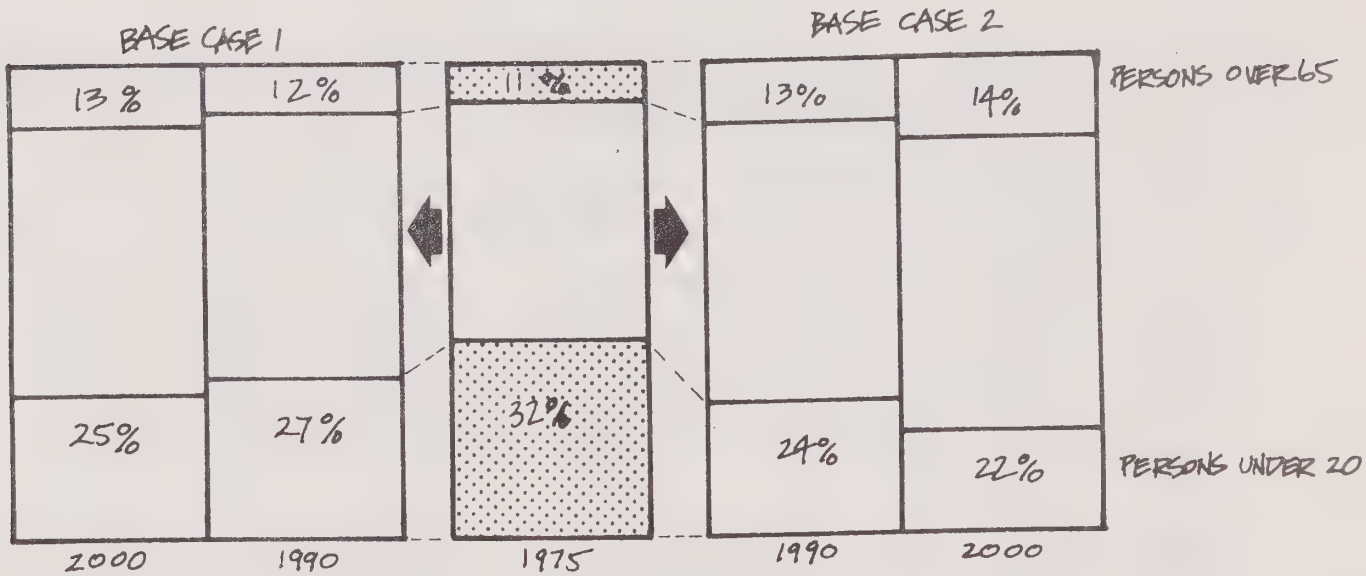


FIGURE 10

PROJECTED CHANGES IN THE PROPORTIONS OF YOUTH AND AGED, 1990 AND 2000



In both Base Cases, deaths continue rising throughout the projection period, due to the growing total population, as well as the increasing median age. The difference between births and deaths--natural increase--declines.

Annual net migration increases throughout the projection period in Base Case 1 and falls throughout the projection period in Base Case 2.

The percentage of population under 20 falls, reflecting lowered fertility rates, and this percentage is thus lower for Base Case 2 than for Base Case 1. The percentage of total population over age 65 increases under both Base Cases, since the present population is weighted with a large number of middle-aged people who migrated to the region during the 1960's. The age structure shifts, most important to health resource planning, are detailed in Table 15.

The historical reduction in household size is projected to continue. The average household size has fallen considerably, especially since 1970. The projected decline in household size is at a lower average household size because of its lower fertility assumption.

TABLE 15
AGE AND SEX DISTRIBUTION OF REGIONAL POPULATION

1975						
AGE GROUP	MALES	FEMALES	TOTAL	PERCENT MALE	PERCENT FEMALE	PERCENT TOTAL
0- 4	172753.	163289.	336042.	7.2	6.7	7.0
5- 9	186601.	177667.	364268.	7.8	7.3	7.5
10-14	218296.	205449.	423745.	9.1	8.4	8.8
15-19	222977.	209859.	432836.	9.3	8.6	9.0
20-24	214685.	200971.	415656.	9.0	8.2	8.6
25-29	188225.	213144.	401369.	7.9	8.7	8.3
30-34	183875.	180907.	364782.	7.7	7.4	7.6
35-39	150778.	144366.	295144.	6.3	5.9	6.1
40-44	132206.	128746.	260952.	5.5	5.3	5.4
45-49	136298.	140200.	276498.	5.7	5.7	5.7
50-54	142934.	147001.	289935.	6.0	6.0	6.0
55-59	121528.	125303.	246831.	5.1	5.1	5.1
60-64	100318.	108225.	208543.	4.2	4.4	4.3
65-69	79353.	88968.	168321.	3.3	3.6	3.5
70-74	55891.	72152.	128043.	2.3	3.0	2.7
75-79	39673.	57769.	97442.	1.7	2.4	2.0
80-84	25590.	40949.	66539.	1.1	1.7	1.4
85+	18567.	33638.	52205.	0.8	1.4	1.1
REGIONAL TOTALS	2390548.	2438603.	4829151.	100.0	100.0	100.0

TABLE 15 (continued)

BASE CASE 1, 1990

AGE GROUP	MALES	FEMALES	TOTAL	PERCENT MALE	PERCENT FEMALE	PERCENT TOTAL
0- 4	193652.	184656.	378307.	7.0	6.5	6.7
5- 9	196766.	189072.	385838.	7.1	6.6	6.9
10-14	192295.	185038.	377332.	7.0	6.5	6.7
15-19	187201.	178134.	365335.	6.8	6.2	6.5
20-24	215404.	203348.	418753.	7.8	7.1	7.4
25-29	242085.	238115.	480199.	8.8	8.3	8.5
30-34	250498.	243522.	494020.	9.1	8.5	8.8
35-39	226577.	222688.	449265.	8.2	7.8	8.0
40-44	200451.	224338.	424789.	7.3	7.8	7.6
45-49	185674.	185534.	371208.	6.7	6.5	6.6
50-54	147099.	144894.	291994.	5.3	5.1	5.2
55-59	122305.	124486.	246791.	4.4	4.4	4.4
60-64	115316.	129378.	244694.	4.2	4.5	4.4
65-69	106290.	128280.	234560.	3.8	4.5	4.2
70-74	77523.	102050.	179573.	2.8	3.6	3.2
75-79	51356.	77218.	128574.	1.9	2.7	2.3
80-84	29342.	49730.	79072.	1.1	1.7	1.4
85+	23944.	49226.	73170.	0.9	1.7	1.3
TOTAL	2763763.	2859701.	5623464.	100.0	100.0	100.0

TABLE 15 (continued)

BASE CASE 1, 2000

AGE GROUP	MALES	FEMALES	TOTAL	PERCENT MALE	PERCENT FEMALE	PERCENT TOTAL
0- 4	187290.	178553.	365843.	6.2	5.7	5.9
5- 9	193291.	185890.	379181.	6.4	5.9	6.2
10-14	204635.	196984.	401619.	6.8	6.3	6.5
15-19	210214.	202189.	412403.	7.0	6.4	6.7
20-24	218437.	209694.	428131.	7.2	6.7	7.0
25-29	210331.	211708.	422039.	7.0	6.8	6.9
30-34	231913.	228182.	460095.	7.7	7.3	7.5
35-39	259456.	251805.	511261.	8.6	8.0	8.3
40-44	255019.	250098.	505117.	8.5	8.0	8.2
45-49	226052.	225242.	451294.	7.5	7.2	7.3
50-54	196554.	222454.	419009.	6.5	7.1	6.8
55-59	173464.	179614.	353078.	5.8	5.7	5.7
60-64	129366.	136676.	266041.	4.3	4.4	4.3
65-69	98648.	113666.	212314.	3.3	3.6	3.5
70-74	83118.	112026.	195144.	2.8	3.6	3.2
75-79	66283.	100335.	166617.	2.2	3.2	2.7
80-84	37827.	65580.	103407.	1.3	2.1	1.7
85+	31614.	65299.	96912.	1.0	2.1	1.6
TOTAL	3013503.	3135989.	6149492.	100.0	100.0	100.0

TABLE 15 (continued)

BASE CASE 2, 1990

AGE GROUP	MALES	FEMALES	TOTAL	PERCENT MALE	PERCENT FEMALE	PERCENT TOTAL
0- 4	156862.	149536.	306397.	6.1	5.6	5.9
5- 9	156199.	149759.	305958.	6.1	5.6	5.8
10-14	139324.	133897.	273221.	5.4	5.0	5.2
15-19	181752.	171649.	353401.	7.1	6.4	6.8
20-24	210161.	193235.	403395.	8.2	7.3	7.7
25-29	230875.	224317.	455192.	9.0	8.4	8.7
30-34	236336.	230772.	467108.	9.2	8.7	8.9
35-39	215423.	213802.	429225.	8.4	8.0	8.2
40-44	193394.	218446.	411839.	7.5	8.2	7.9
45-49	181027.	181471.	362498.	7.0	6.8	6.9
50-54	143535.	141890.	285425.	5.6	5.3	5.5
55-59	120274.	122740.	243014.	4.7	4.6	4.6
60-64	114338.	128244.	242583.	4.5	4.8	4.6
65-69	106478.	127634.	234112.	4.1	4.8	4.5
70-74	77893.	101406.	179299.	3.0	3.8	3.4
75-79	51702.	77022.	128724.	2.0	2.9	2.5
80-84	29499.	49703.	79201.	1.1	1.9	1.5
85+	23944.	48990.	72934.	0.9	1.8	1.4
TOTAL	2569010.	2664506.	5233516.	100.0	100.0	100.0

TABLE 15 (continued)

BASE CASE 2, 2000

AGE GROUP	MALES	FEMALES	TOTAL	PERCENT MALE	PERCENT FEMALE	PERCENT TOTAL
0- 4	134716.	128367.	263083.	5.1	4.7	4.9
5- 9	148903.	142516.	291419.	5.7	5.2	5.4
10-14	159261.	152567.	311828.	6.1	5.6	5.8
15-19	161287.	152948.	314234.	6.2	5.6	5.9
20-24	158684.	143115.	301798.	6.1	5.2	5.6
25-29	187054.	182573.	369626.	7.1	6.7	6.9
30-34	202590.	198818.	401409.	7.7	7.3	7.5
35-39	232950.	226713.	459663.	8.9	8.3	8.6
40-44	232777.	229848.	462626.	8.9	8.4	8.6
45-49	208768.	210784.	419552.	8.0	7.7	7.8
50-54	184766.	212660.	397427.	7.0	7.8	7.4
55-59	166504.	173406.	339910.	6.4	6.3	6.3
60-64	125250.	132508.	257757.	4.8	4.8	4.8
65-69	97484.	111143.	208628.	3.7	4.1	3.9
70-74	83214.	110155.	193369.	3.2	4.0	3.6
75-79	66858.	99520.	166378.	2.6	3.6	3.1
80-84	38253.	65266.	103519.	1.5	2.4	1.9
85+	31743.	64723.	96466.	1.2	2.4	1.8
TOTAL	2621057.	2737626.	5358683.	100.0	100.0	100.0

Summary

In summary, the entire subsection on regional population projections describes the complete process:

1. selecting assumptions
2. collecting demographic data
3. preparing demographic parameters
4. running the APPLE system to produce initial projections
5. reviewing the projections
6. revising assumptions and producing final projections
7. analyzing the effects of different assumptions and final results.

Particular attention is given to health related issues, such as the changing age/sex structure of the population and fertility trends. The San Francisco Bay Area is used to illustrate the process.

3. SOCIAL AREA ANALYSIS

Previously population projections for small subareas have been a major problem in health planning. The failure of past approaches has been the arbitrary definition of the subarea. The approach used here considers small area projections primarily as a problem in allocation of future regional population to rationally defined subareas. Before the approach can be used to allocate regional population to small areas we must know the social, economic, and demographic characteristics and their spatial distribution within the region which assist in formulation of a rational subarea description. The technique which allows us to understand the social structure of a region and the spatial pattern of the social fabric is termed social area analysis. The results of the analysis are used to define social areas, or communities, and are the basis of subarea projections.

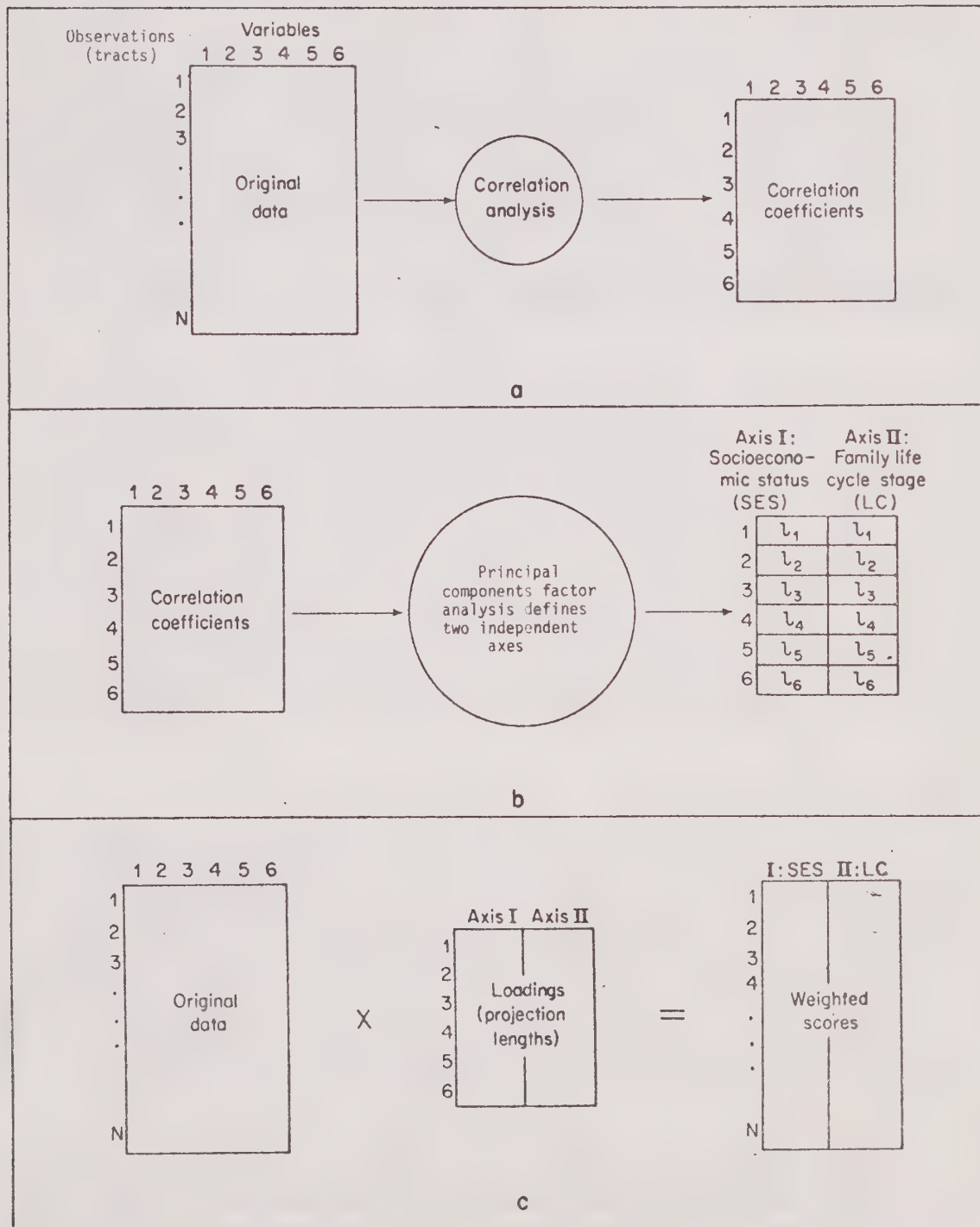
In general, social area analysis, or more correctly, factorial ecology, is the reduction of a large number of socioeconomic variables through factor analytic methods, an examination of the nature of the dimensions, the derivation of measures which index the new dimensions, and the production of profiles of each observation based on the new indices. This process of reducing a large number of descriptive variables to a few indicators is a key step in the project. The dimensions produced allows the planner to understand the pattern of the region's social fabric. The simple indices for each observation are the basis of describing communities within the region. To be effective for use by health planners the social area analysis must have a wide range of socioeconomic variables, and must reduce these to the fewest dimensions that are meaningful for demographic and economic analysis.

The Factor Analysis Method

Factor analysis is a statistical process with two distinct stages. Primarily every factor analysis is a collapsing of a space of intercorrelated variables onto a smaller number of basic dimensions or "composite variables." In a large set of social and economic variables, many tell related stories and are therefore highly correlated. Factor analysis describes each original variable in terms of the factors or latent dimensions of variation. Secondly, factor analysis includes a "rotation" process by which the underlying factors are transformed mathematically to have the maximum intrinsic meaning. The rotation commonly used summarizes mutually exclusive subsets of the original variables. Such a rotation is termed "simple structure."

The factor analysis computational steps are: (1) generation of correlation coefficients, (2) extraction of the latent dimensions or factors, and (3) the calculation of weighted scores. The procedure reveals the latent data structure and reduces the original data to a smaller set of weighted scores or "factor scores." Figure 11 outlines the procedure for a two factor example for six variables. The social area analysis concept demands independent and additive factors. Factorial ecology demands a "closed" model. That is, we

FIGURE 11



Reduction of an original data matrix to a matrix of weighted scores.

hypothesize that the explanation for all variables is contained within the data set. There are no explanatory elements outside of the model. Principal components analysis is "closed." Other factor methods, such as principal axes, are open, and hence allow a unique error term to be carried through the analysis. The principal components factor analysis is a well-established method which meets these demands and therefore its implementation described below.

The procedure begins with a matrix X of order n by m , of n observations and m variables. The next step is to form a matrix Z , again of order n by m in which the original variables have been normalized if necessary and then expressed in the standardized form of zero mean and unit variance. Then form R , the m by m matrix of zero-order correlations among the variables based on Z . The second part of the procedure is the principal components factor analysis. We wish to express each variable of Z , in terms of the factor vectors F . For each j -th variable, then,

$$z_j = a_{j1} F_1 + a_{j2} F_2 + \dots + a_{jm} F_m \quad (43)$$

The factor analysis uses R , the correlation matrix. The communality of each variable, h_j^2 , is approximated by computing the coefficient of determination resulting from the regression of variable j on the factors.

The principal components factor analysis produces a matrix A of order m by r , such that $R = AA^T$ and $A^T A = L$. A^T is the transpose of A . Each element (or factor loading) a_{jk} of A is the correlation coefficient of variable j with factor k , there being r dimensions of variation underlying the original m variables. L is a diagonal matrix containing the eigenvalues associated with each factor. An eigenvalue expresses that portion of the total common variation accounted for by each factor. Since L is diagonal, the cross products of the factors are zero, i.e., they are not correlated, so that each dimension expresses an independent, additive part of the original whole given by the m variables. That part of the original m variables which constitutes each factor may be determined by examining the factor loadings. The one or more variables with high magnitude loadings can be used as an index for the dimension and all of its related variables. Obtaining the best matrix of loading for interpretation is done with second part of the factor analysis process, called rotation.

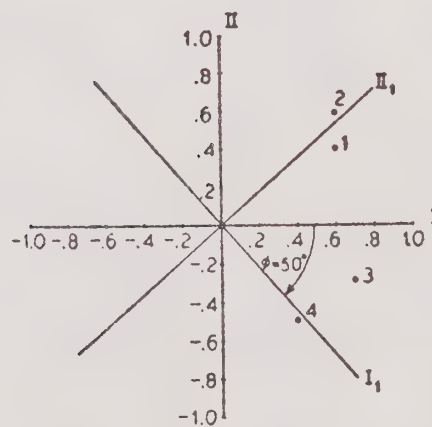
Rotation affects the interpretation of the factors and not the solution. Varimax rotation to simple structure for columns of A is the appropriate choice for the analysis. There are many rotation methods available, but varimax attempts to identify each variable with only one factor, which is consistent with the social area analysis concept. The usefulness of rotation can be demonstrated by a simple example. Begin with a four variable factor matrix of Table 16. Figure 12 is the graphical representation of the matrix.

TABLE 16 FACTOR MATRIX

Variable	Factor		h^2
	I	II	
1	.6	.4	.52
2	.6	.6	.72
3	.7	-.3	.58
4	.4	-.5	.41

FIGURE 12

GRAPHICAL REPRESENTATION
OF TABLE 16 AND ROTATION



A clockwise rotation is done using the relationships:

$$a'_{j1} = a_{j1} \cos \theta - a_{j2} \sin \theta \quad (44)$$

$$a'_{j2} = a_{j2} \sin \theta + a_{j1} \cos \theta \quad (45)$$

If the axes are rotated 50 degrees, the new solution is in Table 17.

A comparison of the original and rotated solutions shows a great improvement in interpretation. Previously factor I had three variables with high loadings. Variable three loaded high on both. The rotated solution is simple to understand. Each factor has two variables closely associated with it. There is less "overlap" of variables. In actual computation with several factors, many possible rotations are attempted before a final structure is selected.

The final part of the factor analysis is the computation of the weighted scores. The original n by m data matrix X is post-multiplied by the m by r loading matrix A to produce the initial weighted score matrix S of size n by r . S is scaled to zero mean and unit variance in its final form. Each S_{ik} is the factor score of observation i on factor k . Thus the original matrix is reduced to the smaller form.

Factor analysis should not be used as a black box. The researcher must specify several parameters that control the factoring process. The analyst must use his judgment while the analysis is performed several times to extract the maximum amount of information. Since the social area analysis concept demands a parsimonious factor structure, two types of decisions are needed: (1) the number of factors, (2) the variables used in the model.

The number of factors extracted is controlled by the maximum number of factors derived and the eigenvalue cutoff. If unconstrained, principal components factor analysis tends to generate as many factors as initial variables. The vast majority of these factors are, however, the unique portions of the variance of the variables. Since these minor factors account for less variance than one variable alone, their eigenvalues are less than one. The eigenvalue cutoff of 1.0 is the initial parameter value. After the first pass over the data, the analyst may request fewer factors to aid interpretation by specifying the maximum number of factors desired. For example, the variance explained may drop suddenly for the final factor. In that case the number of factors requested should be reduced and the program rerun.

TABLE 17 ROTATED FACTOR LOADINGS

Test	Factor		h^2
	I ₁	II ₂	
1	.08	.72	.52
2	-.07	.85	.73
3	.68	.34	.58
4	.64	-.02	.41

The variables used in the analysis is a second control area. If a variable has a low communality in the first pass, its variance is unique and it literally has nothing in common with the entire data set. Generally the analyst should delete this variable before the next analysis. The factor loadings should be carefully examined. There are three types of factors: (1) general factors containing high loadings for all variables, (2) group factors having high loadings for two or more (but not all) variables, (3) unique factors where only one variable is present. The latter, generally, should be deleted from the analysis. The analyst should work with the factor analysis until the solution meets the goals of social area analysis described earlier.

Factor Analysis Computer Programs

There are many factor analysis routines available. Some of the more popular, easy to use programs are: (1) Statistical Package for the Social Sciences (SPSS) developed at the University of Chicago, (2) Princeton's statistical package, called PSTAT, and the common Bio-Medical set developed at University of California, Los Angeles, BMDP. The programs can be accessed through many colleges or any university, and many major commercial computer service bureaus.

Data

The factorial ecology approach has considerable data requirements concerning both depth and breadth of information. Social indicators detailed in fine crosstabulations act as excellent input to the factor analysis. The data should be universally reported for the region in the geographical units to be studied.

These two criteria effect the entire process of variable selection, choice of data sources, selection of observations, and the way that the data are processed. Each of these steps is described in detail below.

Selection of Variables

Many discussions of the use of factor analysis as a method to describe regions are available (a bibliography follows this subsection). Even superficial study of previous analyses reveals that there exists a common data set that provides meaningful results while minimizing data collection and processing work. In this study the suggested variable list of Rees was adapted to our available data.

The accessibility of data sources in part determines the choice of variables. For this analysis Fourth Count Censuses of Population and Housing were chosen as the main data base. Magnetic type versions of these files were obtained for the nine county region. Detailed racial breakdowns for the San Francisco-Oakland SMSA were obtained from a National Cancer Institute study. More discussion of these sources follows. Table 18 is the list of variables and the tabulation numbers used to extract the information from the population and housing files.

TABLE 18
SOCIAL AREA ANALYSIS VARIABLES

Name of Variable	Mnemonic	File/Record Type/Tabulation
<u>Demographic Variables</u>		
% Population Under 18	% POP < 18 yrs	P(1)/17
% Population Under 5	% POP < 5 yrs	P(1)/17
% Population Greater than 65	% POP > 65 yrs	P(1)/17
Mean Age of Population	MEAN AGE	P(1)/17
Average Family Size	AVG. FAMILY S	P(1)/18
Persons Per Household	PERSON/HH	H/7
% Households with Female Head	% HH FEMALE H	P(1)/18
<u>Racial and Ethnic Characteristics</u>		
% Hispanic	% CHICANO	P(4)/17
% Black	% BLACK	P(3)/17
% Chinese	% CHINESE	Cancer Institute
% Indian	% INDIAN	Cancer Institute
% Japanese	% JAPANESE	Cancer Institute
% White	% WHITE	P(2)/17
% Other Races	% OTHER RACES	P(1,2,3)/17
% Population Native Born of Foreign or Mixed Parentage	% OF MIXED	P(1)/22
% Population Native Born of Native Parentage	% OF NATIVE	P(1)/21
<u>Income Characteristics</u>		
Mean Family Income	MEAN FM INCOME	P(1)/75
% Population with Income Over \$15,000/year	% INC > \$15,000	
% Population with Income Under \$5,000/year	% INC < \$ 5,000	P(1)/75
% Population with Income Deficit	INC < 125 % POOR	P(1)/16
Nonwage Income Per Recipient	NON-WAGE/PCAP	P(1)/16,98
<u>Occupational Characteristics</u>		
% Blue Collar	% BLUE COLLAR	P(1)/68
% Clerical and Sales	% CLERIC + SALE	P(1)/68
% Craftsmen and Operatives	% CRAFT + OPR	P(1)/68
% Government	% GOVERNMENT	P(1)/67
% Laborers	% LABORERS	P(1)/68
% Professional and Managerial	% PRO + MNGRL	P(1)/68

TABLE 18 (Continued)

% Self Employed	% SELF-EMP	P(1)/67
% Service Occupations	% SERVICES	P(1)/68
% White Collar	% WHITE COLLAR	P(1)/68
<u>Educational Characteristics</u>		
Mean School Years Completed	MEAN SCHL YR	P(1)/42
% Population Over 25 with Education Past High School	% >25 W HS	P(1)/42
<u>Mobility Characteristics</u>		
% Moved Since 1965	% MOVED-1965	H/10
% Moved Since 1960	% MOVED 1960	H/10
<u>Housing Characteristics</u>		
% Housing Units That Are Trailers	% DU-TRAILERS	H/50
Units With Less Than 1.01 Persons Per Room	% DU > 1.01/RM	H/42
% Units with all Plumbing	% ALL PLUMB	H/59
% Housing Units With Heating	% DU W HEAT	H/59
Housing Value/Income Ratio	HOME \$/INC	H/1, P(1)/1
Mean Home Value	MEAN HOME \$	H/52
Mean Rent	MEAN RENT	H/54
Vacancy Rate-Owned Housing Units	VAC RAT-OWN	H/35,36
Vacancy Rate-Rental Housing Units	VAC RAT-RNT	H/35,36
% Housing Built Before 1940	% HU < 1940	H/8
% Housing Built Before 1950	% HU < 1950	H/8
% Housing Units Built Since 1960	% HU > 1960	H/8
% Housing Units in 1 Unit Structures	% HU 1 UNIT	H/9
% Housing Units in 2 Unit Structures	% HU 2 UNIT	H/9
% Housing Structures With 3 or More Units	% HU 3+ UNIT	H/9
% Owner Occupied Housing	% OWN OCCPD	H/7,8
% Population in Group Quarters	% GROUP Q	P(1)/51
Households With 2 or More Cars	% 2+ CARS	H/17

TABLE 18 (Continued)

Employment Characteristics

% Commuting by Auto	% CMMT-AUTO	P(1)/36
% Commuting by Bus	% CMMT-BUS	P(1)/36
% Commuting by Foot	% CMMT-FOOT	P(1)/36
% Commuting by Rail	% CMMT-RAIL	P(1)/36
% Unemployed	% UNEMP	P(1)/56
Employee/Population Ratio	EMPLOY/POP	P(1)/55
% Males Over 65 in the Labor Force	% MALE > 65 LF	P(1)/17,55
% Women Married with Children in Labor Force	% MRS + KID LF	P(1)/57
% Women Over 14 in Labor Force	% FEMALE > 14 LF	P(1)/17,54

The census tract is the simplest building block of each of these levels. Zip codes provide a unit of analysis that is also useful to health planners, but were determined to be too large to be useful in social area analysis. In addition, the Fourth Count census and the detailed racial characteristics are reported at the tract level. For these reasons, the census tract as defined in 1970 was chosen to be the geographic unit of observation.

Data Processing

It is strongly advised that the analyst consider using a consultant to prepare the needed data base. Processing the large census files to produce the variables needed for the social area analysis will be very expensive unless the analyst is experienced in census tape use. Universities and many colleges have the census materials and the experienced staff to quickly process the tapes for a reasonable fee. Councils of Governments, such as ABAG, or other regional planning agencies can also act as census data agents. Private firms also offer census data. The planner/analyst is encouraged to purchase his variables and skip to the factor analysis results section below. For the braver souls, instructions to process the tapes follows.

Fourth Count census data were obtained from Lawrence Berkeley Laboratory for each of the nine counties. Each county's data for both the Population and Housing files had to be checked for missing data. The result of this processing was regional population and housing files with the same census tract record format for every census tract. Countless further problems were encountered because of idiosyncratic differences between file structures. In solving these problems experimentation provided the best results.

Only after initial processing of the data did certain data errors appear. The tract records had been arranged in ascending tract sequence. To streamline an already lengthy record, the tract records were written in binary format with only two geographic identifiers. (The tract number and county code were sufficient for this analysis.)

The data obtained from the National Cancer Institute required minimal attention. For the five counties in the San Francisco/Oakland SMSA the racial distribution was determined and written to a file. The tract records were then sorted in the same sequence as the Fourth Count data.

Data Processor Design

Before the desired values could be computed for the social area analysis, an understanding of the organization of the census data was required. The 1970 Census Users Guide, Part II was of some help in this effort. The Population file details the tabulations for Total, White, Black, and Spanish-Surname. Except for the non-SMSA racial distribution variables,

the total record type was used. Below is indicated the least troublesome path in the extraction of the desired data set.

1. Determine the variables to be considered in the social area analysis.
2. Associate with each of these variables the tabulation number that will provide the desired information. For the purpose of these directions consider the variable "% income over \$15,000 per year". Reference to the Census User's Guide indicates that the Population File, tabulation 75 "Family Income" contains needed figures.
3. Parallel to 1 and 2 above the process of assigning variables to tabs may be accomplished. Each tabulation contains a specified number of variables, for example Population file tabulation 75 contains 15 data items. But each tract record contains a series of variables' values, with no relationship to the tabulation numbers. Cumulative variable counts should be computed and indicated in the Census User's Guide. For example, Population file tabulation 75 contains variables 876 through 890 ("padding" spaces, as indicated in the Census User's Guide, are not counted as variables).
4. Associate with each desired variable, i.e. "% income over \$15,000", the data items that will be involved in computation. For the example variable, data items 888, 889, and 890 provide counts of families with incomes over \$15,000 per year.
5. Write a processing program to compute the desired values. The data processing program must provide other functions as well. Suppressed data, usually indicated by a data item value of -1., must be set to zero. Before any divisions may be performed, the divisors must be checked for zero values. Certain errors on the census file itself become evident in the values obtained. For instance, percentages greater than 100.00 were detected and screened. The Fortran source language used to process the sixty variables used in the analysis required 300 lines of code.

Factor Analysis Results

The factor analysis was run three times to develop a useful description of the social structure. Analysts will find that this is generally the minimum necessary. The first pass checks for out of range data values and unique factor problems. In the second pass the number of factors is chosen and unique factor variables are eliminated. The third analysis is the basis of the social area study. This example is a general guide to doing a factor analysis.

The First Pass

The first pass at the data employed all sixty-one variables. The first step in analyzing results is always a careful examination of the univariate measurements (see Table 19). The mean, standard deviation, minimum and maximum values outline each variable's distribution. The skewness measures the shift of the distribution away from the mean. The kurtosis is an indicator of how peaked or flat the distribution is. The first pass showed that the large databank was very clean. The minimum and maximum values were correct for the variables. For example, none of the percentage variables exceeded 100.0 per cent. The means were reasonable and conformed to prior expectations. Several variables, such as persons per household, displayed strong skewness or kurtosis problems. Factor analytic methods require variables with normal distributions. Normally, the analyst would replace a variable with a poor distribution with the logarithmic transformation of the variable. However, to keep the analysis as straightforward as possible, transformations which would have removed the problems will not be used. The effects of distributional problems will be seen later in the subsection.

The parameters which control a factor analysis were kept as loose as possible. The only limit for the first pass was an eigenvalue cutoff of 1.00. Other controls, such as the maximum number of factors, were set to high values. To study the factor structure carefully as it evolves the solutions for five through the cutoff dimension number were requested. Experience has shown that solutions of less than five factors for large, complex data sets, are not very useful.

The first pass produced fourteen factors, accounting for 76.4% of the total variance in the final solution. The solution of that many factors is too messy for a social area analysis. The relative strength of each factor must be examined. See Table 20. A scree test reveals breaks in the factor strengths. The first two factors are strong. Factor III is about half the strength of the first factor. The relative strengths diminish slowly then, and taper off at factor X. The remaining four factors contribute little to the solution. For the second pass a ten-factor solution is the upper limit.

The five-factor solution showed that the analysis was developing correctly (Table 21). Factor I was predominately a socioeconomic dimension, but contained a few variables that measure the stage in life cycle. Factor II was almost entirely occupational structure variables. Females in the labor force and related variables formed the next dimension. A clear Black versus White dimension emerged as factor IV. Mobility measures form the final factor. The racial measurements for Chinese, Japanese and others did not enter the solution significantly. Chinese did not emerge as a separate dimension until the six-factor solution. The solutions were too constrained. Simply reducing the number of factors desired for the next pass is not enough. The poorly performing variables must be removed.

TABLE 19
SECOND PASS
UNIVARIATE STATISTICS

VARIABLE	NAME	MEAN	ERROR	N	DEVIATION	ERROR	SKEW	KURTOSIS	MIN	MAX
1	%BLACK	9.073	0.650	1042	20.351	0.968	2.865	7.428	0.0	97.38
2	%CHICANO	12.349	0.344	1042	11.089	0.521	2.320	7.188	0.0	87.44
3	%CHINESE	2.022	0.237	1042	7.637	1.060	8.305	78.344	0.0	96.00
4	%INDIAN	0.266	0.014	1042	0.463	0.033	3.606	19.667	0.0	4.51
5	%JAPANESE	0.713	0.043	1042	1.394	0.161	5.798	53.299	0.0	18.86
6	%WHITE	82.460	0.767	1042	24.753	0.891	-2.098	3.396	0.0	100.00
7	%OTHER RACES	6.643	0.288	1042	9.303	0.923	5.466	39.046	0.0	98.53
8	%CMMT-AUTO	72.164	0.720	1042	23.234	0.777	-1.794	2.656	0.0	99.91
9	%CMMT-BUS	9.672	0.401	1042	12.929	0.477	1.853	3.677	0.0	92.60
10	%CMMT-FOOT	5.833	0.278	1042	8.966	0.674	4.122	21.579	0.0	92.86
11	%CMMT-RAIL	0.702	0.071	1042	2.288	0.620	13.780	304.359	0.0	54.90
12	%BLUE COLLAR	26.470	0.380	1042	12.280	0.292	0.197	0.364	0.0	88.58
13	%CLERIC+SALE	26.443	0.254	1042	8.194	0.279	-0.075	2.837	0.0	71.43
14	%CRAFT+OPR	22.560	0.336	1042	10.859	0.268	0.213	0.547	0.0	88.58
15	%GOVERNMENT	19.645	0.326	1042	10.533	0.642	2.650	13.497	0.0	100.00
16	%LABORERS	4.588	0.132	1042	4.253	0.501	5.341	55.859	0.0	65.22
17	%PRO+MNGRL	25.564	0.430	1042	13.883	0.330	0.713	0.350	0.0	85.72
18	%SELF-EMP	7.203	0.143	1042	4.610	0.194	1.741	5.365	0.0	38.60
19	%SERVICES	12.167	0.197	1042	6.359	0.366	2.162	11.842	0.0	73.69
20	%UNEMP	6.527	0.219	1042	7.078	1.191	9.281	116.044	0.0	100.00
21	%WHITE COLLR	52.003	0.534	1042	17.251	0.360	-0.247	-0.187	0.0	100.00
22	%MOVED-1965	57.825	0.518	1042	16.712	0.497	-0.349	1.684	0.0	100.00
23	%INC>\$15000	25.833	0.517	1042	16.701	0.394	0.887	0.320	0.0	84.87
24	%INC<\$5000	10.477	0.231	1042	7.467	0.328	1.867	6.049	0.0	64.62
25	%2+CARS	40.982	0.660	1042	21.309	0.355	-0.051	-0.844	0.0	100.00
26	%DU-TRAILERS	1.785	0.190	1042	6.136	0.687	6.121	50.220	0.0	80.07
27	%DU>1.01/RM	6.084	0.162	1042	5.215	0.195	1.719	3.801	0.0	33.34
28	%ALL PLUMB	96.177	0.435	1042	14.046	1.329	-5.852	35.296	0.0	100.00
29	%MALE>65 LF	24.279	0.463	1042	14.952	0.564	1.260	3.937	0.0	100.00
30	%MRS+KID LF	37.721	0.364	1042	11.757	0.437	0.006	3.751	0.0	100.00
31	%OF MIXED	17.538	0.204	1042	6.599	0.190	0.208	1.457	0.0	51.79
32	%OF NATIVE	71.113	0.485	1042	15.665	0.713	-2.084	6.630	0.0	100.09
33	%DU W HEAT	98.953	0.224	1042	7.245	1.455	-12.525	166.017	0.0	99.99
34	%>25 W HS	64.611	0.537	1042	17.343	0.422	-0.662	0.466	0.0	100.00
35	EMPLOY/POP	0.393	0.003	1042	0.098	0.004	-1.287	4.143	0.0	0.67
36	HOME \$/INC	1.229	0.020	1042	0.643	0.012	-0.560	-0.649	0.0	2.60
37	INC<125%POOR	0.076	0.002	1042	0.068	0.004	2.780	12.778	0.0	0.65
38	MEAN FM INC	10293.047	275.346	1042	8888.188	2695.180	16.094	381.245	0.0	231809.25
39	MEAN HOME \$	21973.848	369.867	1042	11939.316	268.358	0.089	0.106	0.0	58869.90
40	MEAN RENT	128.190	1.514	1042	48.858	1.330	-0.331	1.090	0.0	315.70

TABLE 19 (Continued)

VARIABLE	NAME	MEAN	ERROR	N	DEVIATION	ERROR	SKEW	KURTOSIS	MIN	MAX
41	MEAN SCHL YR	11.592	0.056	1042	1.811	0.094	-1.662	9.279	0.0	16.30
42	NONWAGE/PCAP	639.299	18.429	1042	594.889	48.592	3.001	25.809	0.0	7562.40
43	VAC RAT-OWN	0.012	0.001	1042	0.026	0.004	8.161	91.134	0.0	0.38
44	VAC RAT-RNT	0.049	0.002	1042	0.061	0.007	6.084	58.173	0.0	0.84
45	%HU <1940	28.053	0.875	1042	28.251	0.514	0.803	-0.620	0.0	100.00
46	%HU <1950	14.567	0.417	1042	13.452	0.558	1.781	5.163	0.0	100.00
47	%HU >1960	30.147	0.791	1042	25.544	0.590	1.030	0.223	0.0	100.00
48	%HU 1 UNIT	64.973	0.921	1042	29.720	0.544	-0.741	-0.604	0.0	100.00
49	%HU 2 UNIT	5.770	0.228	1042	7.354	0.391	2.430	9.788	0.0	72.03
50	%HU 3+ UNIT	25.923	0.822	1042	26.525	0.634	1.143	0.382	0.0	100.00
51	%HH FEMALE H	21.193	0.352	1042	11.371	0.401	0.995	3.176	0.0	100.00
52	%OWN OCCPD	54.004	0.829	1042	26.761	0.434	-0.400	-0.901	0.0	98.47
53	%POP <18 YRS	30.993	0.345	1042	11.142	0.264	-0.415	0.338	0.0	66.47
54	%POP <5 YRS	9.341	0.123	1042	3.970	0.118	0.482	1.705	0.0	30.36
55	%POP >65 YRS	9.511	0.228	1042	7.350	0.471	2.529	15.103	0.0	76.95
56	%GROUP Q	3.092	0.326	1042	10.532	1.175	6.653	49.872	0.0	98.53
57	%FEMALE>14LF	42.614	0.310	1042	9.994	0.362	-0.870	3.457	0.0	74.67
58	AVG FAMILY S	2.620	0.029	1042	0.918	0.022	-0.550	0.368	0.0	5.15 ⁰⁰
59	MEAN AGE	32.329	0.215	1042	6.956	0.297	0.135	5.613	0.0	80.35
60	PERSON/HH	3.115	0.062	1042	1.955	0.524	14.876	288.446	0.0	46.19
61	%MOVED<1960	22.103	0.343	1042	11.055	0.254	0.282	0.188	0.0	70.22

Table 20
FIRST PASS - 14 FACTOR STRUCTURE

Factor	% Variation Explained	Cumulative % Variation Explained
1	15.2	15.2
2	11.6	26.7
3	8.8	35.5
4	7.1	42.6
5	6.8	49.4
6	4.7	54.2
7	3.3	57.5
8	3.2	60.7
9	3.1	63.7
10	3.0	66.7
11	2.8	69.6
12	2.5	72.1
13	2.3	74.4
14	2.0	76.4

TABLE 21
FIRST PASS
5 FACTOR SOLUTION

CASCADING MATRIX OF 5 ROTATED FACTORS		% OF 61 VARS CUMULATIVE % COMMUNALITY	17.1 17.1 1	12.3 29.4 2	9.8 39.2 3	8.4 47.6 4	6.5 54.1 5
NO.	NAME						
48	%HU 1 UNIT	0.892	0.894	-0.010	0.191	-0.064	-0.227
25	%2+CARS	0.892	0.851	-0.224	0.193	-0.227	0.171
52	%OWN OCCPD	0.818	0.836	-0.106	0.076	-0.250	-0.199
58	AVG FAMILY S	0.852	0.815	0.163	0.313	0.127	0.218
50	%HU 3+ UNIT	0.832	-0.811	-0.062	0.270	0.115	0.289
36	HOME \$/INC	0.751	0.717	-0.208	0.299	-0.175	-0.273
53	%POP <18 YRS	0.756	0.704	0.300	0.270	0.201	0.239
23	%INC>\$15000	0.819	0.649	-0.537	0.199	-0.262	-0.034
45	%HU <1940	0.680	-0.648	0.008	0.198	0.253	-0.397
9	%CMMT-BUS	0.591	-0.641	-0.067	0.257	0.289	-0.162
8	%CMMT-AUTO	0.586	0.621	0.109	0.397	-0.175	0.010
51	%HH FEMALE H	0.647	-0.615	-0.032	0.321	0.356	-0.197
39	MEAN HOME \$	0.742	0.597	-0.461	0.319	-0.233	-0.133
10	%CMMT-FOOT	0.344	-0.583	-0.042	-0.033	0.028	-0.018
7	%OTHER RACES	0.306	-0.523	0.134	0.054	-0.107	0.012
55	%POP >65 YRS	0.529	-0.501	-0.114	0.148	-0.014	-0.493
3	%CHINESE	0.261	-0.469	0.049	0.149	-0.124	-0.027
32	%OF NATIVE	0.364	0.447	-0.075	0.036	0.302	0.257
38	MEAN FM INC	0.248	0.427	-0.224	0.022	-0.121	0.001
49	%HU 2 UNIT	0.381	-0.418	0.118	0.204	0.358	-0.150
60	PERSON/HH	0.202	0.307	-0.010	0.068	0.270	0.174
4	%INDIAN	0.156	-0.304	0.204	0.006	0.145	0.029
5	%JAPANESE	0.172	-0.261	-0.212	0.219	0.033	-0.101
12	%BLUE COLLAR	0.815	0.196	0.868	0.155	-0.011	-0.010
14	%CRAFT+OPR	0.775	0.217	0.830	0.168	-0.099	0.042
17	%PRO+MNGRL	0.779	0.245	-0.802	0.181	-0.206	-0.006
21	%WHITE COLLR	0.887	0.105	-0.786	0.461	-0.215	-0.001
34	%>25 W HS	0.859	0.314	-0.766	0.296	-0.254	0.147
41	MEAN SCHL YR	0.744	0.278	-0.713	0.360	-0.130	0.106
27	%DU>1.01/FM	0.678	-0.018	0.687	0.140	0.418	0.104
2	%CHICANO	0.517	0.026	0.684	0.192	-0.105	0.031
16	%LABORERS	0.345	0.061	0.495	-0.022	0.258	-0.170
54	%POP <5 YRS	0.681	0.391	0.447	0.323	0.223	0.417
11	%CMMT-RAIL	0.066	0.105	-0.219	0.047	-0.066	0.013
26	%DU-TRAILERS	0.071	0.059	0.199	-0.034	-0.070	0.150
28	%ALL PLUMB	0.836	0.376	-0.041	0.810	0.188	0.035
57	%FEMALE>14LF	0.710	-0.208	-0.082	0.803	-0.029	0.117
33	%DU W HEAT	0.635	0.218	-0.023	0.736	0.210	-0.018
13	%CLERIC+SALE	0.579	-0.176	-0.328	0.655	-0.111	0.006
30	%MRS+KID LF	0.423	-0.138	0.153	0.586	0.191	-0.026
	TOTAL	33.002	10.422	7.526	5.954	5.129	3.971

TABLE 21 (Continued)

CASCADING MATRIX OF 5 ROTATED FACTORS			17.1	12.3	9.8	8.4	6.5
NO. NAME			17.1	29.4	39.2	47.6	54.1
COMMUNALITY			1	2	3	4	5
40	MEAN RENT	0.587	0.256	-0.374	0.548	-0.274	0.076
35	EMPLOY/POP	0.553	-0.196	-0.196	0.536	-0.422	-0.101
56	%GROUP Q	0.350	-0.188	-0.152	-0.520	0.110	0.096
29	%MALE>65 LF	0.197	0.147	-0.230	0.286	-0.174	-0.099
1	%BLACK	0.665	-0.121	0.168	-0.041	0.785	-0.068
37	INC<125%POOR	0.718	-0.062	0.452	0.084	0.708	0.048
24	%INC<\$5000	0.709	-0.116	0.452	0.134	0.687	-0.036
6	%WHITE	0.524	0.282	-0.220	0.070	-0.622	0.072
19	%SERVICES	0.488	-0.273	0.307	0.063	0.562	0.012
31	%OF MIXED	0.630	-0.182	-0.041	0.395	-0.533	-0.393
15	%GOVERNMENT	0.327	0.011	-0.232	-0.068	0.516	-0.048
43	VAC RAT-CWN	0.112	0.001	-0.007	-0.008	0.305	0.138
20	%UNEMP	0.157	-0.148	0.228	-0.146	0.246	-0.030
42	NONWAGE/PCAP	0.123	0.094	-0.067	-0.200	-0.227	0.135
61	%MOVED<1960	0.776	0.129	0.014	0.245	0.014	-0.836
47	%HU >1960	0.744	0.374	-0.127	0.134	-0.171	0.735
22	%MOVED-1965	0.750	-0.115	0.008	0.480	0.159	0.693
59	MEAN AGE	0.557	-0.415	-0.216	0.206	-0.144	-0.524
46	%HU <1950	0.345	0.085	0.147	0.147	0.273	-0.469
18	%SELF-EMP	0.384	0.228	-0.388	0.120	-0.050	-0.406
44	VAC RAT-RAT	0.084	-0.023	-0.012	0.102	0.107	0.247
TOTAL		33.002	10.422	7.526	5.954	5.129	3.971

Returning to the final fourteen factor solution, variables with less than half of the variance explained by the solution should be removed from the analysis. See Table 22. Only two variables had communalities less than 0.500 and were removed for the second pass: vacancy rate owner occupied units and vacancy rates of renter occupied units. Although the Japanese variable had a communality of 0.462 it is considered an important demographic indicator and must be kept. Other variables distorted the analysis in more subtle manners.

1. The commuter variables, percent by bus and percent by foot, loaded mildly on several dimensions.
2. The occupational structure variables were discovered to be linear transformations of each other and hence always formed a separate dimension. The census variables used in the percent variables of blue collar, crafts and operatives, government, laborers, professional and managerial, self-employed, and white collar were removed.
3. The variable nonwage income per recipient formed a unique factor and told nothing more about social space, and was removed.
4. Percent Indian is scattered across several dimensions and hence was useless as a demographic indicator.
5. When several percentage variables can be summed to 100% they become a linear transformation of each other and hence form their own dimension. The percent units in two unit structures was removed so it could not be combined with percent in one unit or percent in three or more unit structures.
6. The variables percent unemployed and mean family income formed a separate dimension that revealed nothing. The former variable was removed.
7. The unique factor of percent dwelling units that are mobile homes must be deleted.
8. Males over 65 in the labor force stands alone as an esoteric, unique factor and was deleted for the second pass.
9. Those commuting by rail formed a unique factor for the pre-BART 1970 census, and was removed.
10. The variable percent housing units built before 1950 was scattered across several dimensions and was removed.

Therefore, only forty-four variables were input to the second pass factor analysis.

TABLE 22

FIRST PASS
14 FACTOR SOLUTION

CASCADING MATRIX OF 14 ROTATED FACTORS			% OF 61 VARS CUMULATIVE % COMMUNALITY											
NO.	NAME		15.2 15.2 1	11.6 26.7 2	8.8 35.5 3	7.1 42.6 4	6.8 49.4 5	4.7 54.2 6	3.3 57.5 7	3.2 60.7 8	3.1 63.7 9	3.0 66.7 10	2.8 69.6 11	2.5 72.1 12
48	%HU 1 UNIT	0.931	-0.940	0.047	0.060	0.017	-0.039	-0.132	-0.003	-0.057	0.095	0.006	-0.018	0.009
52	%OWN OCCPD	0.896	-0.917	-0.087	-0.034	0.037	-0.109	-0.077	-0.065	0.078	-0.060	0.039	0.006	-0.094
36	HOMF \$/INC	0.790	-0.819	-0.149	0.179	-0.099	-0.095	-0.141	-0.039	-0.014	-0.083	-0.013	0.063	0.112
25	%2+CARS	0.928	-0.813	-0.188	0.147	0.297	-0.215	-0.123	-0.031	0.022	0.104	-0.192	0.003	-0.063
50	%HU 3+ UNIT	0.892	0.811	-0.072	0.384	-0.043	0.015	0.148	0.112	-0.013	-0.065	-0.168	-0.031	0.096
58	AVG FAMILY S	0.935	-0.752	0.195	0.218	0.432	0.034	-0.056	0.223	-0.076	0.163	-0.008	0.037	0.018
23	%INC>\$15000	0.884	-0.724	-0.501	0.110	0.090	-0.226	0.003	0.040	0.071	0.003	-0.058	0.040	0.087
39	MEAN HOME %	0.807	-0.664	-0.413	0.230	-0.003	-0.195	-0.133	-0.015	0.001	-0.148	-0.071	-0.012	0.198
8	%CMMT-AUTC	0.735	-0.601	0.164	0.319	0.015	-0.086	-0.288	-0.035	0.007	-0.146	-0.243	0.243	-0.110
53	%POP <18 YRS	0.873	-0.587	0.312	0.142	0.472	0.131	-0.094	0.228	-0.143	0.005	-0.076	0.150	0.128
51	%HH FEMALE H	0.775	0.556	-0.033	0.343	-0.327	0.264	-0.047	0.115	-0.259	-0.196	0.172	-0.005	0.088
10	%CMMT-FOOT	0.718	0.484	-0.043	-0.021	-0.216	0.009	0.452	-0.094	-0.111	0.229	-0.225	0.057	0.077
9	%CMMT-BUS	0.797	0.461	-0.065	0.271	-0.235	0.209	0.252	0.284	0.091	0.021	0.390	-0.095	0.257
12	%BLUE COLLAR	0.902	-0.153	0.904	0.095	0.012	0.047	-0.084	-0.035	0.076	0.005	-0.091	0.120	-0.098
14	%CRAFT+OPR	0.883	-0.189	0.865	0.120	0.043	-0.017	-0.057	-0.048	0.125	0.002	-0.103	0.164	-0.129
17	%PRO+MNGRL	0.854	-0.292	-0.768	0.129	-0.049	-0.240	-0.036	0.019	-0.087	-0.001	-0.125	0.219	0.094
21	%WHITE COLLR	0.932	-0.198	-0.727	0.419	-0.086	-0.300	-0.020	0.072	-0.072	-0.012	0.031	0.247	0.125
2	%CHICANO	0.732	0.035	0.723	0.082	0.093	-0.301	-0.043	0.126	-0.158	0.000	0.068	-0.004	0.203
34	%>25 W HS	0.892	-0.317	-0.715	0.292	0.107	-0.281	-0.115	0.014	0.113	0.041	-0.019	0.226	0.020
27	%DU>1.01/RM	0.786	0.048	0.658	0.062	0.238	0.355	0.218	0.228	-0.187	0.009	0.032	-0.031	0.125
41	MEAN SCHL YR	0.835	-0.268	-0.634	0.349	0.017	-0.194	-0.185	0.075	0.117	0.112	-0.032	0.302	0.033
57	%FEMALE>14LF	0.879	0.091	-0.040	0.877	0.007	0.072	0.137	-0.209	0.034	-0.082	-0.000	0.130	0.046
28	%ALL PLUMP	0.899	-0.395	0.031	0.757	0.023	0.008	-0.191	0.234	-0.225	0.028	0.073	-0.034	-0.068
33	%DU W HEAT	0.782	-0.251	0.048	0.689	-0.090	-0.014	-0.113	0.168	-0.370	0.120	-0.055	-0.107	-0.025
30	%MRS+KID LF	0.721	-0.028	0.184	0.669	-0.113	0.330	0.241	-0.175	0.058	0.096	-0.064	-0.084	-0.026
13	%CLERIC+SALE	0.703	0.063	-0.262	0.653	-0.098	-0.230	0.016	0.116	-0.008	-0.025	0.263	0.157	0.105
22	%MOVED-1965	0.868	0.262	0.013	0.576	0.386	-0.102	-0.051	0.369	-0.063	0.155	-0.260	-0.069	-0.250
40	MEAN RENT	0.692	-0.345	-0.334	0.557	0.074	-0.224	-0.031	-0.096	0.040	-0.173	-0.091	-0.027	0.186
35	EMPLOY/POP	0.809	0.031	-0.144	0.526	-0.260	-0.166	0.072	-0.203	0.366	-0.299	-0.019	0.362	-0.016
59	MEAN AGE	0.855	0.202	-0.134	0.127	-0.836	-0.120	0.023	0.073	0.091	-0.023	-0.069	0.189	-0.068
55	%POP >65 YRS	0.824	0.313	-0.074	0.080	-0.789	-0.104	0.038	0.141	-0.192	-0.054	-0.062	-0.052	-0.107
61	%MOVED<1960	0.807	-0.376	0.084	0.103	-0.632	0.145	0.030	-0.229	-0.141	-0.104	0.284	0.023	0.230
47	%HL >1960	0.835	-0.211	-0.171	0.183	0.597	-0.256	-0.070	0.229	0.027	0.005	-0.350	0.039	-0.324
45	%HL <1940	0.795	0.438	0.032	0.158	-0.548	0.144	0.226	0.240	-0.061	0.016	0.187	-0.184	0.205
54	%POP <5 YRS	0.795	-0.212	0.446	0.243	0.538	0.062	-0.148	0.332	-0.155	-0.051	0.002	0.181	0.050
1	%BLACK	0.847	0.156	0.144	0.030	-0.006	0.862	-0.094	0.134	0.048	0.013	0.034	-0.106	0.116
6	%WHITE	0.746	-0.223	-0.154	0.043	0.007	-0.763	-0.249	-0.086	0.026	0.044	-0.025	0.048	-0.060
31	%OF MIXED	0.801	-0.065	0.021	0.210	-0.437	-0.598	0.275	0.061	-0.061	-0.131	0.209	0.147	0.166
37	INC<125%POOR	0.844	0.169	0.424	-0.005	0.121	0.534	-0.028	0.448	-0.325	-0.013	0.076	0.059	0.030
24	%INC<\$5000	0.843	0.197	0.435	0.040	-0.014	0.512	-0.049	0.427	-0.364	-0.016	0.070	0.085	-0.040
	TOTAL	46.603	9.245	7.046	5.353	4.356	4.165	2.886	2.020	1.935	1.865	1.847	1.735	
	TOTAL	1.499												

TABLE 22 (Continued)

CASCADING MATRIX OF 14 ROTATED FACTORS		% OF 61 VARS	2.3	2.0
NO.	NAME	CUMULATIVE % COMMUNALITY	74.4 13	76.4 14
48	%HU 1 UNIT	0.931	-0.095	-0.033
52	%OWN OCCPD	0.896	-0.002	0.059
36	HOME \$/INC	0.790	0.030	-0.010
25	%2+CARS	0.928	-0.058	-0.065
50	%HU 3+ UNIT	0.892	0.020	-0.044
58	AVG FAMILY S	0.935	-0.055	-0.079
23	%INC>\$15000	0.884	-0.103	-0.080
39	MEAN HOME %	0.807	-0.054	-0.136
8	%CMMT-AUTC	0.735	0.007	0.047
53	%POP <18 YRS	0.873	-0.209	0.013
51	%HH FEMALE H	0.775	0.111	-0.021
10	%CMMT-FOOT	0.718	-0.306	0.053
9	%CMMT-BUS	0.797	0.074	0.151
12	%BLUE COLLAR	0.902	-0.043	0.057
14	%CRAFT+OPR	0.883	0.077	0.042
17	%PRO+MNGRL	0.854	-0.142	0.018
21	%WHITE COLLR	0.932	-0.047	-0.022
2	%CHICANO	0.732	-0.107	0.046
34	%>25 W HS	0.892	-0.145	-0.066
27	%DU>1.01/RM	0.786	-0.104	-0.037
41	MEAN SCHL YR	0.835	-0.195	-0.066
57	%FEMALE>14LF	0.879	-0.070	0.032
28	%ALL PLUMP	0.899	-0.063	-0.104
33	%DU W HEAT	0.782	-0.163	0.002
30	%MRS+KID LF	0.721	0.013	0.060
13	%CLERIC+SALE	0.703	0.130	-0.074
22	%MOVED-1965	0.868	-0.052	-0.063
40	MEAN RENT	0.692	0.088	-0.046
35	EMPLOY/POP	0.809	-0.111	-0.024
59	MEAN AGE	0.855	-0.091	-0.005
55	%POP >65 YRS	0.824	0.013	0.045
41	%MOVED<1960	0.807	-0.030	-0.088
47	%HU >1960	0.835	-0.005	0.136
45	%HU <1940	0.795	-0.097	0.152
54	%POP <5 YRS	0.795	-0.055	-0.024
1	%BLACK	0.847	0.001	0.053
6	%WHITE	0.746	0.036	-0.087
31	%OF MIXED	0.801	-0.101	-0.012
37	INC<125%PCOR	0.849	-0.074	-0.137
24	%INC<\$5000	0.843	-0.059	-0.130
	TOTAL	46.603	1.427	1.225

TABLE 22 (Continued)

CASCADINO MATRIX OF 14 ROTATED FACTORS NO. NAME	% OF 61 VARS CUMULATIVE % COMMUNALITY	15.2	11.6	8.8	7.1	6.8	4.7	3.3	3.2	3.1	3.0	2.8	2.5
		15.2 1	26.7 2	35.5 3	42.6 4	49.4 5	54.2 6	57.5 7	60.7 8	63.7 9	66.7 10	69.6 11	72.1 12
19 %SERVICES	0.642	0.249	0.324	0.104	-0.088	0.510	0.180	0.174	-0.034	0.326	-0.016	0.064	-0.031
3 %CHINESE	0.854	0.162	-0.012	0.098	-0.083	0.062	0.894	-0.002	0.042	-0.011	-0.033	0.068	0.036
7 %OTHER RACES	0.845	0.262	0.053	0.013	-0.011	0.038	0.871	-0.005	-0.007	-0.000	0.079	0.027	-0.057
32 %OF NATIVE	0.716	-0.201	-0.002	0.091	0.185	0.247	-0.519	0.020	0.095	0.261	-0.298	0.280	-0.066
43 VAC RAT-OWN	0.418	-0.002	-0.017	-0.053	-0.013	0.195	-0.001	0.579	0.131	0.025	-0.121	-0.039	-0.049
42 NONWAGE/PCAP	0.686	-0.090	-0.051	-0.177	0.146	-0.048	-0.026	0.111	0.747	-0.048	0.112	0.151	0.071
4 %INDIAN	0.543	0.300	0.260	0.082	-0.164	0.068	-0.097	0.194	0.378	0.166	-0.083	-0.219	0.235
60 PERSON/HM	0.706	-0.254	0.072	0.128	0.167	0.008	0.020	0.128	-0.031	0.739	0.065	-0.097	-0.041
56 %GROUP Q	0.690	0.303	-0.106	-0.394	0.008	-0.031	-0.034	-0.142	0.103	0.626	-0.015	0.062	0.054
15 %GOVERNMENT	0.702	0.040	-0.180	0.005	-0.078	0.395	-0.131	-0.055	-0.229	0.520	0.047	0.282	0.015
44 VAC RAT-RNT	0.401	0.042	-0.024	0.137	0.029	0.106	0.007	0.154	-0.032	-0.069	-0.557	-0.119	0.061
49 %HU 2 UNIT	0.571	0.368	0.121	0.200	-0.116	0.190	0.014	0.252	-0.116	-0.027	0.452	-0.108	0.169
5 %JAPANESE	0.462	0.135	-0.213	0.275	-0.066	0.107	0.136	-0.039	0.230	-0.008	0.452	-0.029	-0.054
20 %UNEMP	0.564	0.195	0.191	-0.067	-0.024	0.157	-0.072	0.021	-0.101	-0.015	-0.070	-0.656	0.033
38 MEAN FM INC	0.602	-0.480	-0.250	0.031	0.124	-0.128	0.030	0.012	-0.023	-0.034	-0.039	-0.515	-0.027
26 %DU-TRAILERS	0.722	-0.011	0.141	-0.014	0.020	-0.050	-0.015	0.047	-0.082	0.001	0.019	0.012	-0.829
29 %MALE>65 LF	0.650	-0.187	-0.204	0.245	0.016	-0.090	0.040	-0.157	0.158	-0.077	0.153	-0.048	0.007
16 %LABORERS	0.666	0.015	0.500	-0.135	-0.086	0.163	-0.067	0.103	-0.214	0.025	0.001	-0.080	-0.056
18 %SELF-EMP	0.716	-0.350	-0.371	-0.077	-0.367	-0.097	0.069	0.192	-0.223	-0.083	0.036	0.057	-0.102
11 %CMMT-RAIL	0.695	-0.048	-0.189	0.024	0.036	-0.089	-0.008	0.003	0.054	-0.004	-0.061	-0.017	0.025
46 %HU <1950	0.626	-0.116	0.216	0.117	-0.321	0.283	-0.160	-0.156	-0.109	0.090	0.284	0.019	0.039
TOTAL	46.603	9.245	7.046	5.353	4.356	4.165	2.886	2.020	1.935	1.865	1.847	1.735	
TOTAL	1.499												

TABLE 22 (Continued)

CASCADING MATRIX OF 14 ROTATED FACTORS		% OF 61 VARS	2.3	2.0
NO.	NAME	CUMULATIVE % COMMUNALITY	74.4 13	76.4 14
19	%SERVICES	0.642	0.109	-0.097
3	%CHINESE	0.854	0.008	-0.000
7	%OTHER RACES	0.845	-0.039	0.041
32	%OF NATIVE	0.716	-0.230	0.021
43	VAC RAT-OWN	0.418	0.049	0.049
42	NONWAGE/PCAP	0.686	0.002	-0.082
4	%INDIAN	0.543	-0.033	0.134
60	PERSON/HH	0.706	0.036	-0.106
56	%GROUP Q	0.690	-0.007	0.029
15	%GOVERNMENT	0.702	0.041	0.280
44	VAC RAT-RNT	0.401	0.089	0.036
49	%HU 2 UNIT	0.571	0.088	0.036
5	%JAPANESE	0.462	-0.098	0.129
20	%UNEMP	0.564	-0.087	0.004
38	MEAN FM INC	0.602	0.041	-0.063
26	%DU-TRAILERS	0.722	-0.025	0.034
29	%MALE>65 LF	0.650	-0.646	-0.064
16	%LABORERS	0.666	-0.539	0.047
18	%SELF-EMP	0.716	-0.433	-0.068
11	%CMMT-RAIL	0.695	-0.048	-0.798
46	%HU <1950	0.626	0.084	-0.457
	TOTAL	46.603	1.427	1.225

The final inspection for the first pass is a careful review of the factor scores for the fourteen-factor solution. Tracts with extreme values are suspected of having incorrect data or from poor observational units because of special conditions. The latter case applied to sixteen tracts that were ships in harbor at the time of the census. These tracts were removed (17999, 22699, 60699, 252199, 314299, 315099, 320099, 378099, 380099, 401799, 401999, 403299, 427299, 427399, 427499, 427599). The second pass used 1,042 tracts as input.

The Second Pass

The second pass univariate statistics reveal some variation over the pattern observed in the first pass. Although the change is slight some of the variables which were altered are: percent white collar, percent units with heat, mean family income, and nonwage income per recipient.

The second pass stopped with ten factors as requested. The five-factor solution in the first pass was too tight, so printout was requested beginning with the six-factor solution. Since the poor observations and variables were removed, the factor solutions may now be examined in detail. The relative strengths of each factor for the second pass solutions are displayed in Table 23.

The six-factor solution explains 68.2% of total variance. See Table 24. Factor I, strongly represents the socioeconomic status factor as in pass one. Percentage variance explained declines strongly after this factor. The second factor describes a race (Black)/poverty relationship. The variable, percent service occupations, loads significantly on this factor. Percent White and percent born of mixed parentage load significantly as negative. Factor III combines mobility and weak indicators of stage in family life cycle. Mean family size and percent population under 18 years, both stage in family life cycle indicators, load on factors I and III. Factor IV combines working wives' families with the renter population and those working in clerical and sales occupations. After factor IV there is another decline in percentage of variance explained to 8.2% for factor V, which shares heavy poverty loadings with factor II. Besides poverty, poor education, and overcrowding, the Hispanics are represented by factor V. Factor six is dominated by variables percent other races, percent Chinese, and percent native born of native parents (inverse relationship). This last factor accounts for 7.1% of total variance. Superficial analysis of this six-factor solution leads to the conclusion that too much of the variable interaction is stifled in this tight space.

The seven-factor solution Table 25 explains 71.7% of the total variance. The first factor is similar to factor I for the six-factor solution, however in this space it accounts for 20.0% of the total common variance. The first factor, socioeconomic status, remains dominant. Percentage variance explained dips quickly to 10.4% for factor II with heavy loadings in mobility, age and family variables, indicating the strength of the family life cycle factor in this solution. Factor III describes the working women with children dimension (factor IV of the previous factor solution).

Table 23
SECOND PASS SOLUTIONS

PERCENT VARIATION EXPLAINED					
FACTOR SOLUTION					
Factor	6	7	8	9	10
I	19.7	20.0	19.3	19.7	18.8
II	12.0	10.4	13.0	12.4	11.0
III	10.8	10.2	10.6	10.1	9.4
IV	10.6	10.1	7.7	8.6	7.2
V	8.2	9.7	6.9	6.5	7.0
VI	7.1	7.1	6.7	6.5	6.3
VII	-	4.2	6.6	6.3	6.3
VIII	-	-	4.2	4.2	6.2
IX	-	-	-	3.4	4.1
X	-	-	-	-	3.6
Total	68.2	71.7	74.9	77.5	80.0

TABLE 24
SECOND PASS
6 FACTOR SOLUTION

CASCADED 6	MATRIX OF ROTATED FACTORS NO. NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	14.7	12.0	10.8	10.6	9.2	7.1
			19.7	31.6	42.4	53.0	61.1	68.2
			1	2	3	4	5	6
48	%HU 1 UNIT	0.896	0.921	-0.052	0.001	0.040	-0.068	-0.200
52	%OWN OCCPD	0.874	0.907	-0.188	0.013	0.036	0.040	-0.113
50	%HU 3+ UNIT	0.850	-0.860	0.042	-0.067	0.264	0.072	0.173
25	%2+CARS	0.903	0.799	-0.272	-0.359	0.109	0.136	-0.177
36	HOME \$/INC	0.768	0.787	-0.174	0.137	0.243	0.084	-0.185
58	AVG FAMILY \$	0.890	0.756	0.138	-0.459	0.104	-0.246	-0.129
23	%INC>\$15000	0.811	0.714	-0.352	-0.106	0.183	0.359	-0.055
39	MEAN HOME \$	0.713	0.644	-0.321	0.026	0.312	0.268	-0.159
51	%HH FEMALE H	0.759	-0.642	0.302	0.381	0.326	-0.003	-0.061
53	%POP <18 YRS	0.792	0.606	0.221	-0.438	0.043	-0.396	-0.161
8	%CMT-AUTO	0.542	0.531	-0.130	-0.092	0.303	-0.181	-0.332
38	MEAN FM INC	0.295	0.463	-0.209	-0.111	-0.066	0.136	0.038
1	%BLACK	0.792	-0.123	0.873	0.054	-0.038	0.067	-0.079
6	%WHITE	0.703	0.182	-0.775	-0.064	0.078	-0.004	-0.242
37	INC<125%POOR	0.751	-0.196	0.689	-0.082	-0.064	-0.456	-0.136
24	%INC<\$5000	0.739	-0.256	0.664	0.019	-0.028	-0.450	-0.168
19	%SERVICES	0.538	-0.282	0.652	-0.013	-0.094	-0.098	0.121
31	%OF MIXED	0.735	-0.001	-0.556	0.460	0.289	-0.243	0.268
61	%MOVED<1960	0.782	0.356	0.116	0.792	0.111	-0.004	0.046
47	%HU >1960	0.727	0.150	-0.306	-0.763	0.105	0.023	-0.135
59	MEAN AGE	0.687	-0.326	-0.166	0.687	0.212	0.179	-0.066
22	%MOVED-1965	0.774	-0.416	-0.024	-0.687	0.338	-0.076	-0.096
55	%POP >65 YRS	0.677	-0.460	-0.131	0.663	0.039	0.053	-0.067
45	%HU <1940	0.634	-0.493	0.231	0.530	0.097	-0.001	0.217
60	PERSON/HH	0.220	0.223	0.206	-0.323	-0.102	0.108	-0.031
57	%FEMALE>14LF	0.751	-0.171	0.065	-0.053	0.793	0.111	0.269
35	EMPLOY/POP	0.776	-0.071	-0.274	0.199	0.781	0.172	0.132
28	%ALL PLUMB	0.734	0.316	0.040	-0.134	0.705	-0.137	-0.303
13	%CLERIC+SALE	0.491	-0.159	-0.247	0.084	0.607	0.161	0.049
40	MEAN RENT	0.563	0.313	-0.326	-0.078	0.546	0.233	0.013
33	%DU W HEAT	0.403	0.130	0.023	-0.010	0.518	-0.140	-0.312
30	%MRS+KID LF	0.522	-0.027	0.415	0.041	0.488	0.011	0.329
56	%GROUP Q	0.424	-0.249	0.058	-0.103	-0.478	0.346	-0.016
5	%JAPANESE	0.249	-0.121	0.077	0.118	0.268	0.266	0.267
2	%CHICANO	0.659	-0.053	-0.056	-0.028	-0.028	-0.807	0.009
27	%DU>1.01/RM	0.793	-0.036	0.566	-0.163	-0.067	-0.636	0.189
54	%POP <5 YRS	0.737	0.183	0.205	-0.528	0.139	-0.577	-0.175
34	%>25 W HS	0.832	0.301	-0.446	-0.140	0.413	0.565	-0.131
41	MEAN SCHL YR	0.752	0.245	-0.323	-0.145	0.474	0.541	-0.221
7	%OTHER RACES	0.790	-0.231	0.084	0.001	0.009	-0.051	0.852
	TOTAL	28.661	8.260	5.030	4.519	4.436	3.428	2.990

TABLE 24 (Continued)

CASCADING MATRIX OF 6 ROTATED FACTORS NO.	NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	19.7	12.0	10.8	10.6	8.2	7.1
			19.7	31.6	42.4	53.0	61.1	68.2
			1	2	3	4	5	6
3	CHINESE	0.764	-0.154	0.069	0.051	0.081	0.029	0.852
32	OF NATIVE	0.572	0.196	0.245	-0.342	0.033	0.170	-0.571
	TOTAL	28.661	8.260	5.030	4.518	4.436	3.428	2.990

TABLE 25
SECOND PASS
7 FACTOR SOLUTION

CASCADING MATRIX OF 7 ROTATED FACTORS NO. NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	20.0	10.4	10.2	10.1	9.7	7.1	4.2
		20.0 1	30.4 2	40.6 3	50.7 4	60.4 5	67.5 6	71.7 7
48 %HU 1 UNIT	0.912	-0.927	0.024	0.059	-0.052	0.074	-0.162	0.082
52 %OWN OCCPD	0.875	-0.911	-0.001	0.016	-0.142	-0.081	-0.115	-0.072
50 %HU 3+ UNIT	0.851	0.858	-0.069	0.260	0.036	-0.072	0.185	-0.041
25 %2+CARS	0.908	-0.810	-0.345	0.130	-0.227	-0.165	-0.165	0.097
36 HOME %/INC	0.768	-0.795	0.133	0.230	-0.132	-0.120	-0.169	-0.078
58 AVG FAMILY S	0.926	-0.754	-0.410	0.152	0.102	0.292	-0.098	0.214
23 %INC>\$15000	0.819	-0.727	-0.093	0.189	-0.283	-0.400	-0.040	0.073
39 MEAN HOME \$	0.713	-0.655	0.017	0.299	-0.248	-0.326	-0.146	-0.080
51 %HH FEMALE H	0.759	0.646	0.390	0.325	0.273	0.045	-0.037	-0.076
53 %POP <18 YRS	0.796	-0.603	-0.437	0.047	0.201	0.415	-0.158	-0.021
8 %CMMT-AUTO	0.580	-0.535	-0.132	0.275	-0.093	0.122	-0.330	-0.261
38 MEAN FM INC	0.295	-0.467	-0.111	-0.069	-0.175	-0.163	0.033	0.044
61 %MOVED<1960	0.782	-0.351	0.781	0.067	0.114	0.005	0.056	-0.166
47 %HU >1960	0.729	-0.161	-0.766	0.130	-0.270	-0.067	-0.135	0.059
59 MEAN AGE	0.720	0.315	0.706	0.226	-0.188	-0.183	-0.035	0.010
55 %POP >65 YRS	0.715	0.451	0.685	0.063	-0.180	-0.045	-0.044	0.048
22 %MOVED-1965	0.807	0.400	-0.637	0.410	-0.061	0.106	-0.056	0.235
45 %HU <1940	0.682	0.493	0.568	0.111	0.167	0.061	0.247	0.106
28 %ALL PLUMB	0.797	-0.336	-0.075	0.761	0.052	0.175	-0.229	0.118
57 %FEMALE>14LF	0.790	0.169	-0.091	0.721	0.131	-0.157	0.295	-0.322
35 EMPLOY/POP	0.829	0.062	0.138	0.698	-0.188	-0.277	0.153	-0.430
13 %CLERIC+SALE	0.514	0.139	0.103	0.617	-0.238	-0.191	0.099	-0.007
33 %DU W HEAT	0.474	-0.149	0.046	0.580	-0.029	0.176	-0.248	0.142
40 MEAN RENT	0.574	-0.325	-0.107	0.506	-0.243	-0.314	0.031	-0.205
30 %MRS+KID LF	0.523	0.034	0.045	0.446	0.427	0.026	0.356	-0.105
1 %BLACK	0.822	0.154	0.064	-0.055	0.883	0.050	-0.094	-0.014
6 %WHITE	0.725	-0.211	-0.066	0.116	-0.776	-0.093	-0.225	0.034
31 %OF MIXED	0.903	-0.028	0.463	0.284	-0.616	0.166	0.312	-0.058
47 INC<125%POOR	0.763	0.207	-0.034	-0.019	0.584	0.530	-0.117	0.162
19 %SERVICES	0.574	0.294	0.051	-0.052	0.579	0.232	0.140	0.270
2 %HICANO	0.669	0.041	-0.032	-0.013	-0.179	0.794	0.028	-0.057
27 %DU>1.01/RM	0.794	0.046	-0.148	-0.067	0.466	0.713	0.147	0.020
34 %>25 W HS	0.839	-0.318	-0.174	0.431	-0.354	-0.616	-0.106	0.076
54 %POP <5 YRS	0.740	-0.186	-0.524	0.155	0.149	0.593	-0.163	-0.040
24 %INC<\$5000	0.756	0.265	0.069	0.020	0.554	0.513	-0.145	0.156
41 MEAN SCHL YR	0.774	-0.263	-0.110	0.511	-0.251	-0.563	-0.183	0.136
7 %OTHER RACES	0.808	0.232	0.011	-0.029	0.054	0.063	0.862	0.062
3 %CHINESE	0.772	0.155	0.054	0.030	0.062	-0.028	0.860	0.008
32 %OF NATIVE	0.580	-0.139	-0.325	0.074	0.283	-0.123	-0.576	0.035
5 %JAPANESE	0.249	0.122	0.124	0.245	0.111	-0.260	0.281	-0.017
TOTAL	50.111	8.413	4.372	4.270	4.253	4.072	2.972	1.753

TABLE 25 (Continued)

CASCADING MATRIX OF 7 ROTATED FACTORS NO.	NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	20.0	10.4	10.2	10.1	9.7	7.1	4.2
			20.0	30.4	40.6	50.7	60.4	67.5	71.7
			1	2	3	4	5	6	7
60	PERSON/HH	0.710	-0.240	-0.167	0.051	0.101	0.046	0.035	0.780
56	GROUP Q	0.595	0.246	-0.007	-0.370	0.010	-0.243	-0.006	0.581
	TOTAL	30.111	8.413	4.372	4.276	4.253	4.072	2.972	1.753

The Black/White/poverty factor ranks fourth. Overcrowding, poverty, and working mother indicators spill into this factor, enriching its interpretation. Factor V is similar to the fifth factor in the six-factor solution. The sixth factor in this solution describes the other races dimension. Factor VII describes persons per household and percent group quarters. The seven-factor solution breaks these variables out of factor I where they were marginally associated in the six-factor solution.

The eight-factor solution, Table 26, accounts for 74.9% of total variance, but the marginal gain acts to split the factors somewhat artificially. The three strongest factors are again socioeconomic status, Black/White, and stage in family life cycle. Group quarters and other races maintain their last and next to last position, respectively. The two factors representing working mothers and Chicano/overcrowding are split into three heavily overlapping factors, considerably muddying the interpretation.

Until the nine-factor solution, Table 27, the variable mean family income experienced very low communality with the factor space (less than .245). In this solution the variable generates its own factor, accounting for 3.4% of total variance. Group quarters and other races hold positions 8 and 7. A useless factor VI appears, with loadings only on percent dwelling units with heat and percent dwelling units with all plumbing.

The ten-factor solution accounts for 80.0% of total variance. See Table 28. The factor added, factor V, is split from the family life cycle factor in the nine factor solution. The new factor, in accounting for 7.0% of total variance, concerns mobility measures: percent moved before 1960, percent moved since 1965, percent new housing.

Given the goals of the study the parsimonious seven factor solution was chosen for further analysis. The third and final pass will seek the seven factor solution. Mean family income was deleted from the factor analysis in the third run because it formed its own factor.

The Third Pass

The third and final pass showed the social structure of the region. The seven factors accounted for 72.8% of the common variance of the forty-one variables. The factor strengths and loadings are displayed in Table 29. The factor scores were mapped to aid in the interpretation of the social space and to understand the spatial implications.

Factor I is a strong dimension describing socioeconomic status as revealed in choice of lifestyle. High status is indicated by the variables with high negative loadings. A high status area has one unit homes, owner occupied, with a high ratio of home value to family income, and high home value. The area also has a large percentage of families earning over \$15,000 a year. They own two or more cars and drive to work. The family size is above average with a higher proportion of members under 18 years old. The typical low status tracts not only have the "reverse" values for the variables mentioned (renter occupied, low income, etc.) but are also

TABLE 26
SECOND PASS
8 FACTOR SOLUTION

CASCADING MATRIX OF 8 ROTATED FACTORS NO.	NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	19.3	13.0	10.6	7.7	6.9	6.7	6.6	4.2
			19.3 1	32.2 2	42.8 3	50.5 4	57.5 5	64.1 6	70.7 7	74.9 8
48	%HU 1 UNIT	0.914	-0.924	-0.074	0.012	-0.033	0.133	0.069	-0.165	0.062
52	%OWN OCCPD	0.875	-0.899	-0.195	-0.006	0.089	0.021	0.049	-0.110	-0.077
50	%HU 3+ UNIT	0.861	0.858	0.044	-0.066	0.093	0.155	-0.205	0.193	-0.077
25	%2+CARS	0.909	-0.786	-0.306	-0.342	0.168	0.126	0.011	-0.167	0.093
36	HOME \$/INC	0.769	-0.782	-0.200	0.128	0.150	0.202	-0.057	-0.151	-0.108
58	AVG FAMILY S	0.927	-0.764	0.133	-0.439	-0.160	0.269	0.042	-0.065	0.147
23	%INC>\$15000	0.838	-0.642	-0.381	-0.094	0.423	0.157	0.020	0.001	0.011
51	%HH FEMALE H	0.794	0.632	0.313	0.377	0.072	0.309	-0.148	0.022	-0.179
39	MEAN HOME \$	0.746	-0.625	-0.332	0.012	0.377	0.256	-0.029	-0.090	-0.167
53	%POP <18 YRS	0.797	-0.619	0.257	-0.465	-0.240	0.146	0.037	-0.144	-0.063
8	%CMMT-AUTO	0.638	-0.536	-0.167	-0.118	-0.168	0.158	-0.246	-0.406	-0.172
38	MEAN FM INC	0.469	-0.443	-0.136	-0.143	0.285	0.078	0.319	0.169	-0.138
1	%BLACK	0.828	0.117	0.871	0.062	0.117	-0.082	-0.117	-0.130	-0.007
6	%WHITE	0.733	-0.173	-0.784	-0.058	-0.039	0.177	0.141	-0.180	0.014
37	INC<125%POOR	0.840	0.170	0.773	-0.087	-0.315	0.251	0.209	0.001	-0.020
24	%INC<\$5000	0.820	0.228	0.739	0.019	-0.322	0.284	0.188	-0.026	-0.022
27	%DU>1.01/RM	0.794	-0.001	0.616	-0.187	-0.581	0.044	-0.018	0.199	-0.004
19	%SERVICES	0.602	0.258	0.603	0.051	-0.176	-0.049	-0.133	0.078	0.336
31	%OF MIXED	0.805	-0.019	-0.568	0.446	-0.262	0.265	-0.091	0.356	-0.092
40	MEAN RENT	0.583	-0.300	-0.360	-0.103	0.329	0.305	-0.308	0.027	-0.236
61	%MOVED<1960	0.785	-0.363	0.083	0.780	-0.019	0.005	-0.122	0.039	-0.134
47	%HU >1960	0.743	-0.133	-0.277	-0.772	0.107	0.164	0.071	-0.097	-0.010
59	MEAN AGE	0.724	0.321	-0.221	0.714	0.112	0.174	-0.110	-0.019	0.010
55	%POP >65 YRS	0.729	0.455	-0.120	0.677	0.038	0.160	0.130	0.050	-0.046
22	%MOVED-1965	0.816	0.412	-0.028	-0.651	-0.004	0.441	-0.101	-0.005	0.133
54	%POP <5 YRS	0.743	-0.204	0.264	-0.564	-0.452	0.281	0.014	-0.125	-0.119
45	%HU <1940	0.694	0.477	0.224	0.553	-0.008	0.134	-0.050	0.247	0.040
2	%SCHICANO	0.763	0.011	-0.037	-0.050	-0.867	0.085	-0.023	-0.004	0.008
34	%>25 W HS	0.850	-0.272	-0.510	-0.157	0.611	0.243	-0.153	-0.083	0.026
41	MEAN SCHL YR	0.775	-0.223	-0.410	-0.090	0.572	0.375	-0.210	-0.170	0.096
5	%JAPANESE	0.253	0.124	0.035	0.129	0.282	0.077	-0.262	0.257	-0.013
28	%ALL PLUMB	0.313	-0.324	0.053	-0.106	0.019	0.778	-0.244	-0.147	-0.038
33	%DU W HEAT	0.491	-0.146	0.005	0.018	-0.021	0.661	-0.099	-0.148	-0.016
13	%CLERIC+SALE	0.516	0.156	-0.304	0.102	0.209	0.466	-0.331	0.121	-0.065
57	%FEMALE>14LF	0.901	0.159	-0.066	-0.056	0.074	0.221	-0.888	0.100	-0.128
30	%MRS+KID LF	0.664	0.003	0.284	0.070	-0.070	0.069	-0.730	0.162	0.099
35	EMPLOY/POP	0.875	0.070	-0.372	0.170	0.167	0.260	-0.719	0.022	-0.299
7	%OTHER RACES	0.823	0.218	0.091	-0.011	-0.051	-0.133	-0.152	0.849	0.060
3	%CHINESE	0.716	0.144	0.064	0.036	0.034	-0.127	-0.221	0.835	0.018
32	%OF NATIVE	0.654	-0.138	0.176	-0.284	0.115	0.027	-0.115	-0.664	0.191
	TOTAL	31.457	8.090	5.444	4.445	3.250	2.911	2.749	2.753	1.744

TABLE 26 (Continued)

CASCADING MATRIX OF S ROTATED FACTORS NO.	NAME	% OF 42 VARS CUMULATIVE % COMMUNALITY	19.3	13.0	10.6	7.7	6.9	5.7	6.6	4.2
			19.3	32.2	42.8	50.5	57.5	64.1	70.7	74.9
			1	2	3	4	5	6	7	8
60	PERSON/HH	0.781	-0.241	0.088	-0.164	-0.034	0.177	0.039	0.020	0.609
56	%GROUP Q	0.715	0.253	-0.039	0.032	0.099	-0.305	0.159	-0.076	0.718
	TOTAL	31.437	8.090	5.444	4.445	3.250	2.911	2.799	2.753	1.744

TABLE 27
SECOND PASS
9 FACTOR SOLUTION

CASCADED MATRIX OF ROTATED FACTORS NO. NAME	1 OF 42 VARS CUMULATIVE COMMUNALITY	14.7 19.7 1	12.4 32.1 2	10.1 42.1 3	8.6 50.7 4	6.5 57.2 5	6.5 63.7 6	6.3 69.9 7	4.2 74.1 8	3.4 77.5 9
48 WHU 1 UNIT	0.414	-0.017	-0.079	0.074	0.004	-0.065	0.146	-0.152	0.060	-0.005
52 *DOWN OCCPD	0.878	-0.890	-0.190	0.058	0.123	-0.051	0.016	-0.094	-0.082	-0.108
50 WHU 3+ UNIT	0.901	0.844	0.056	-0.136	0.053	0.200	0.132	0.187	-0.082	0.116
58 AVS FAMILY S	0.925	-0.318	0.145	-0.390	-0.103	-0.061	0.252	-0.017	0.116	0.024
25 *2+CARS	0.909	-0.800	-0.297	-0.296	0.210	-0.001	0.108	-0.152	0.091	-0.115
36 HOME \$/INC	0.779	-0.775	-0.187	0.171	0.220	0.039	0.102	-0.121	-0.127	-0.004
53 *POP <18 YRS	0.848	-0.697	0.276	-0.422	-0.219	-0.086	0.107	-0.072	-0.114	0.154
23 *INC>\$15000	0.846	-0.672	-0.350	-0.056	0.480	-0.026	0.116	0.024	-0.001	-0.100
51 WHH FEMALE H	0.801	0.664	0.304	0.325	0.051	0.155	0.320	-0.002	-0.172	0.059
39 MEAN HOME \$	0.740	-0.600	-0.312	0.042	0.424	0.038	0.231	-0.086	-0.163	-0.128
8 *CMMT-AUTO	0.639	-0.575	-0.157	-0.096	-0.096	0.226	0.144	-0.374	-0.192	0.130
45 WHU <1940	0.645	0.524	0.201	0.514	-0.032	0.047	0.161	0.277	0.044	0.043
1 *BLACK	0.832	0.134	0.869	0.087	0.033	0.133	-0.069	-0.140	0.014	-0.002
6 *WHITE	0.736	-0.191	-0.781	-0.083	0.030	-0.134	0.176	-0.187	0.012	0.032
37 INC<125%POOR	0.874	0.124	0.772	-0.044	-0.301	-0.254	0.238	0.048	-0.060	0.189
24 *INC<\$5000	0.355	0.185	0.738	0.032	-0.299	-0.238	0.271	0.022	-0.065	0.230
19 *SERVICES	0.606	0.234	0.549	0.047	-0.147	0.104	-0.044	0.108	0.319	0.159
31 *OF MIXED	0.826	-0.015	-0.583	0.415	-0.160	0.044	0.262	0.381	-0.134	0.234
61 *MOVED<1960	0.736	-0.298	0.052	0.812	-0.022	0.123	0.047	0.022	-0.122	-0.042
47 WHU >1960	0.746	-0.196	-0.239	-0.777	0.151	-0.074	0.111	-0.070	-0.030	0.002
22 *MOVED-1960	0.318	0.548	-0.001	-0.704	0.030	0.100	0.404	0.014	0.112	0.117
59 MEAN AGE	0.703	0.330	-0.205	0.666	0.206	0.037	0.139	0.028	-0.036	0.352
55 *POP >65 YRS	0.745	0.484	-0.125	0.621	0.032	-0.163	0.156	0.059	-0.067	0.203
54 *POP <5 YRS	0.772	-0.298	0.276	-0.559	-0.383	-0.057	0.248	-0.064	-0.170	0.223
2 *SCHICANO	0.771	-0.032	-0.085	-0.062	-0.841	0.018	0.142	-0.007	-0.003	0.174
34 *>25 W HS	0.401	-0.246	-0.444	-0.169	0.722	0.107	0.198	-0.022	-0.017	0.127
41 MEAN SCHL YR	0.363	-0.263	-0.336	-0.112	0.708	0.140	0.269	-0.085	0.035	0.256
27 *DU>1.03/RM	0.744	-0.037	0.558	-0.167	-0.605	0.009	0.075	0.211	-0.019	0.068
40 MEAN RENT	0.596	-0.275	-0.346	-0.094	0.353	0.338	0.289	0.008	-0.224	-0.139
57 *FEMALE>14LF	0.409	0.149	-0.054	-0.072	0.047	0.886	0.202	0.106	-0.133	0.122
30 *MRS+KID LF	0.712	0.023	0.264	0.026	-0.114	0.759	0.164	0.140	0.122	-0.034
35 *EMPLOY/POP	0.976	0.069	-0.353	0.145	0.237	0.697	0.223	0.040	-0.318	0.175
5 *JAPANESE	0.254	0.155	0.041	0.126	0.266	0.269	0.066	0.248	-0.007	-0.056
28 *ALL PLUMB	0.840	-0.334	0.053	-0.113	0.071	0.265	0.784	-0.149	-0.043	-0.009
33 *DU W HEAT	0.592	-0.114	-0.017	0.003	-0.020	0.156	0.705	-0.195	0.012	-0.136
13 *CLERIC+SALE	0.525	0.147	-0.278	0.059	0.296	0.298	0.419	0.152	-0.100	0.194
3 *CHINESE	0.840	0.145	0.069	0.047	0.040	0.179	-0.160	0.865	-0.017	0.063
7 *OTHER RACES	0.844	0.223	0.035	-0.004	-0.055	0.126	-0.143	0.865	0.035	0.019
32 *OF NATIVE	0.715	-0.275	0.232	-0.292	0.149	0.047	-0.045	-0.574	0.142	0.326
60 PERSON/HH	0.367	-0.247	0.081	-0.156	-0.041	-0.019	0.209	0.022	0.817	-0.035
TOTAL	32.550	9.266	5.201	4.230	3.601	2.723	2.718	2.625	1.770	1.417

TABLE 27 (Continued)

CASCADING MATRIX OF 4 ROTATED FACTORS		% OF 42 VARS	19.7	12.4	10.1	8.6	6.5	6.5	6.3	4.2	3.4
NO.	NAME	CUMULATIVE % COMMUNALITY	14.7	32.1	42.1	50.7	57.2	63.7	69.9	74.1	77.5
			1	2	3	4	5	6	7	8	9
56	%GROUP Q	0.715	0.246	-0.032	0.013	0.074	-0.163	-0.295	-0.069	0.725	0.017
38	MEAN FM INC	0.760	-0.330	-0.182	-0.091	0.144	-0.163	0.170	0.031	-0.033	-0.729
	TOTAL	32.550	8.265	5.201	4.230	3.601	2.723	2.718	2.025	1.770	1.417

TABLE 28
SECOND PASS
10 FACTOR SOLUTION

CASCADING MATRIX OF 10 ROTATED FACTORS	10 NO.	NAME	1 OF 40 VARS CUMULATIVE COMMUNALITY	10.8 1	11.0 2	9.4 3	7.2 4	7.0 5	6.3 6	6.3 7	6.2 8	4.1 9	3.0 10
46	YHU 1 UNIT	0.901	-0.912	-0.068	0.024	-0.090	0.115	0.162	-0.053	-0.131	0.063	-0.085	
52	DOWN OCCPD	0.895	-0.896	-0.166	0.147	-0.098	0.123	0.020	-0.043	-0.076	-0.076	-0.116	
50	YHU 3+ UNIT	0.862	0.822	0.065	0.077	0.141	-0.249	0.047	0.197	0.175	-0.090	0.080	
25	Y2+CARS	0.970	-0.809	-0.246	0.241	-0.213	-0.219	0.038	0.013	-0.136	0.091	-0.139	
35	HOME \$/INC	0.762	-0.776	-0.156	0.252	0.041	0.179	0.171	0.044	-0.109	-0.120	-0.027	
58	AVG FAMILY S	0.935	-0.769	0.113	-0.035	-0.425	-0.192	0.286	-0.075	-0.007	0.128	0.098	
51	YHH FEMALE H	0.812	0.675	0.284	0.044	0.265	0.223	0.323	0.130	-0.019	-0.167	0.064	
23	2INC>\$15000	0.846	-0.675	-0.390	0.517	-0.058	-0.016	0.080	-0.035	0.022	0.049	-0.147	
8	YCHMT-AUTO	0.707	-0.628	-0.116	-0.085	0.027	-0.154	0.126	0.278	-0.334	-0.267	0.089	
53	YPOP <18 YRS	0.857	-0.629	0.218	-0.143	-0.492	-0.178	0.160	-0.107	-0.069	-0.096	0.261	
39	MEAN HOME \$	0.746	-0.593	-0.271	0.464	-0.008	0.066	0.204	0.028	-0.088	-0.157	-0.164	
45	YHU <1940	0.695	0.510	0.194	-0.068	0.408	0.350	0.160	0.042	0.272	0.041	0.022	
1	YBLACK	0.932	0.165	0.864	-0.044	-0.069	0.144	-0.023	0.102	-0.126	0.004	-0.061	
6	YWHITE	0.746	-0.196	-0.780	0.164	0.021	-0.105	0.159	-0.113	-0.207	0.077	0.015	
37	INC<125YPOOR	0.879	0.162	0.702	-0.304	-0.135	-0.027	0.289	-0.279	0.047	-0.057	0.278	
24	YINC<\$5000	0.855	0.195	0.683	-0.322	-0.000	-0.006	0.308	-0.249	0.027	-0.068	0.292	
31	YOF MIXED	0.935	-0.012	-0.615	-0.062	0.291	0.314	0.250	0.052	0.364	-0.114	0.241	
19	YSERVICES	0.624	0.214	0.548	-0.258	0.085	-0.016	-0.028	0.113	0.129	0.301	0.157	
34	Y>25 W HS	0.910	-0.274	-0.368	0.803	-0.031	-0.140	0.127	0.079	-0.052	0.006	0.063	
41	MEAN SCHL YR	0.878	-0.249	-0.279	0.788	0.042	-0.121	0.193	0.116	-0.113	0.057	0.186	
2	YCHICANO	0.787	0.006	-0.195	-0.753	-0.199	0.039	0.215	0.032	-0.007	0.009	0.306	
27	YDUP>1.01/YM	0.795	0.018	0.494	-0.584	-0.308	-0.016	0.161	-0.008	0.224	-0.017	0.198	
40	MEAN RENT	0.600	-0.245	-0.324	0.429	-0.104	-0.031	0.271	0.315	-0.001	-0.210	-0.149	
13	YCLERIC+SALE	0.601	0.208	-0.297	0.419	0.023	0.091	0.399	0.253	0.113	-0.067	0.211	
5	YJAPANESE	0.422	0.231	-0.006	0.356	-0.181	0.311	0.101	0.187	0.201	0.030	0.019	
59	MEAN AGE	0.698	0.186	-0.112	0.129	0.856	0.224	0.044	0.104	0.052	-0.062	0.179	
55	YPOP >65 YRS	0.897	0.325	-0.042	-0.031	0.854	0.158	0.069	-0.095	0.087	-0.101	0.037	
54	YPOP <5 YRS	0.747	-0.227	0.189	-0.284	-0.522	-0.322	0.303	-0.081	-0.070	-0.152	0.355	
61	YMOVED<1960	0.817	-0.239	0.021	0.002	0.247	0.819	0.095	0.094	0.013	-0.101	-0.001	
22	YMOVED-1965	0.872	0.275	0.035	0.053	-0.094	-0.798	0.345	0.132	0.029	0.090	0.064	
47	YHU >1960	0.621	-0.270	-0.182	0.148	-0.230	-0.797	0.050	-0.034	-0.048	-0.051	-0.063	
28	YALL PLUMB	0.640	-0.322	0.048	0.145	-0.060	-0.104	0.782	0.256	-0.142	-0.035	0.011	
33	YDU W HEAT	0.597	-0.133	-0.015	0.010	0.072	-0.064	0.700	0.144	-0.185	0.010	-0.137	
57	YFEMALE>14LF	0.912	0.190	-0.050	0.139	-0.085	-0.016	0.206	0.862	0.106	-0.124	0.134	
30	YMRS+KID LF	0.762	0.005	0.293	-0.125	0.062	0.039	0.120	0.772	0.179	0.103	-0.053	
35	YEMPLOY/POP	0.877	0.059	-0.314	0.312	0.194	0.067	0.157	0.699	0.042	-0.313	0.124	
3	YCHINESE	0.861	0.127	0.090	0.023	0.116	-0.025	-0.172	0.180	0.883	-0.029	0.026	
7	YOTHER RACES	0.860	0.230	0.077	-0.066	0.017	-0.019	-0.142	0.115	0.872	0.030	0.019	
32	YOF NATIVE	0.719	-0.305	0.278	0.185	-0.080	-0.287	-0.020	0.064	-0.563	0.135	0.284	
60	PERSON/YH	0.808	-0.256	0.080	-0.031	-0.114	-0.124	0.210	-0.019	0.024	0.819	-0.026	
	TOTAL	33.514	7.916	4.629	3.954	3.011	2.951	2.657	2.639	2.599	1.741	1.514	

TABLE 28 (Continued)

CASCADING MATRIX OF		X OF 42 VARS CUMULATIVE COMMUNALITY	18.8	11.0	9.4	7.2	7.0	6.3	6.3	6.2	4.1	3.6
10	ROTATED FACTORS		17.8	22.9	39.3	46.5	53.5	59.8	66.1	72.3	76.4	80.0
NO.	NAME		1	2	3	4	5	6	7	8	9	10
56	SG GROUP C	0.743	0.261	-0.028	0.050	0.012	0.021	-0.309	-0.171	-0.093	0.732	0.044
58	MEAN FM INC	0.770	-0.321	-0.162	0.102	-0.154	-0.035	0.105	-0.107	0.042	-0.041	-0.730
	TOTAL	53.614	7.910	4.629	3.924	3.011	2.951	2.657	2.639	2.599	1.741	1.514

TABLE 29

THIRD PASS

CASCADED MATRIX OF 7 ROTATED FACTORS NO.	NAME	% OF 41 VARS CUMULATIVE % COMMUNALITY	19.5	11.1	10.6	10.2	10.0	7.1	4.3
			14.5	30.6	41.2	51.4	61.4	68.5	72.3
			1	2	3	4	5	6	7
48	%HU 1 UNIT	0.912	-0.931	-0.018	-0.073	0.042	0.042	-0.165	0.090
52	%OWN OCCPD	0.875	-0.905	-0.041	-0.166	-0.005	-0.111	-0.098	-0.062
50	%HU 3+ UNIT	0.857	0.862	-0.026	0.056	0.279	-0.046	0.168	-0.054
36	HOME \$/INC	0.774	-0.806	0.044	-0.151	0.207	-0.151	-0.147	-0.072
25	%2+CARS	0.905	-0.784	-0.374	-0.251	0.117	-0.195	-0.154	0.106
58	AVG FAMILY \$	0.928	-0.760	-0.447	0.081	0.145	0.257	-0.081	0.224
23	%INC>\$15000	0.813	-0.700	-0.115	-0.306	0.170	-0.424	-0.032	0.079
39	MEAN HOME \$	0.707	-0.640	-0.004	-0.264	0.233	-0.348	-0.140	-0.030
51	%HH FEMALE H	0.769	0.613	0.418	0.297	0.338	0.064	-0.049	-0.097
53	%POP <18 YRS	0.799	-0.611	-0.475	0.184	0.042	0.352	-0.135	-0.010
8	%CHMT-AUTO	0.575	-0.553	-0.159	-0.104	0.263	0.091	-0.303	-0.252
47	%HU >1960	0.729	-0.118	-0.766	-0.282	0.138	-0.079	-0.139	0.062
61	%MOVED<1960	0.782	-0.394	0.761	0.113	0.043	-0.002	0.069	-0.167
59	MEAN AGE	0.713	0.280	0.726	-0.172	0.270	-0.171	-0.032	0.003
55	%POP >65 YRS	0.716	0.423	0.709	-0.159	0.067	-0.021	-0.056	0.053
22	%MOVED-1965	0.814	0.425	-0.610	-0.055	0.434	0.110	-0.070	0.227
45	%HU <1940	0.686	0.467	0.591	0.185	0.118	0.084	0.224	0.092
1	%BLACK	0.825	0.129	0.059	0.884	-0.051	0.050	-0.090	-0.021
6	%WHITE	0.725	-0.193	-0.063	-0.730	0.115	-0.046	-0.229	0.037
31	%OF MIXED	0.804	-0.045	0.468	-0.613	0.277	0.169	0.314	-0.058
37	INC<125\$POUR	0.765	0.177	-0.039	0.547	-0.001	0.583	-0.121	0.149
19	%SERVICES	0.575	0.271	0.055	0.588	-0.045	0.237	0.145	0.272
29	%ALL PLUMB	0.805	-0.349	-0.084	0.051	0.767	0.153	-0.227	0.106
57	%FEMALE>14LF	0.784	0.153	-0.080	0.129	0.712	-0.173	0.318	-0.316
35	EMPLOY/POP	0.824	0.046	0.148	-0.136	0.634	-0.240	0.171	-0.428
13	%CLERIC+SALE	0.512	0.131	0.120	-0.255	0.613	-0.196	0.104	-0.012
33	%DU W HEAT	0.444	-0.166	0.046	-0.025	0.589	0.186	-0.253	0.127
40	MEAN RENT	0.572	-0.309	-0.111	-0.255	0.496	-0.330	0.034	-0.209
30	%MRS+KID LF	0.525	0.014	0.045	0.425	0.440	0.015	0.373	-0.100
2	%CHICANO	0.644	0.016	-0.038	-0.172	0.000	0.793	0.032	-0.055
27	%DU>1.01/RM	0.794	0.025	-0.161	0.470	-0.055	0.704	0.201	-0.020
34	%>25 W HS	0.843	-0.295	-0.173	-0.368	0.414	-0.635	-0.094	0.081
41	MEAN SCHL YR	0.777	-0.254	-0.108	-0.267	0.494	-0.595	-0.165	0.141
24	%INC<\$5000	0.758	0.228	0.068	0.569	0.035	0.578	-0.148	0.142
54	%POP <5 YRS	0.737	-0.146	-0.547	0.145	0.155	0.572	-0.110	-0.057
3	%CHINESE	0.773	0.167	0.061	0.055	0.021	-0.021	0.659	0.013
7	%OTHER RACES	0.809	0.244	0.071	0.049	-0.024	0.072	0.658	0.062
32	%OF NATIVE	0.509	-0.200	-0.336	0.290	0.070	-0.145	-0.550	0.076
5	%JAPANESE	0.248	0.110	0.134	0.110	0.238	-0.219	0.282	-0.014
60	PERSON/HH	0.721	-0.252	-0.172	0.093	0.053	0.036	0.039	0.719
	TOTAL	0.852	0.007	4.543	4.344	4.194	4.691	2.909	1.764

TABLE 29 (Continued)

CASCADING MATRIX OF 7 ROTATED FACTORS NO.	NAME	1 OF 41 VARS	19.5	11.1	10.0	10.7	10.0	7.1	4.2
		CUMULATIVE % COMMUNALITY	19.5	30.6	41.2	51.4	61.4	68.5	72.8
			1	2	3	4	5	6	7
56	%GROUP 4	0.607	0.261	0.004	0.013	-0.368	-0.224	-0.004	0.543
	TOTAL	29.852	6.007	4.543	4.344	4.144	4.041	2.903	1.764

marked by a large percentage of three or more unit structures and a high proportion of female head of household. The map of factor I scores (see Map 1) displays the strong spatial components of inner old city decay and the suburban ring of high status communities. The old parts of San Francisco, Berkeley, Oakland, San Jose, etc., all contain the lower status tracts. The slopes and suburbs contain the high status portions of the region.

The stage in family life cycle dimension emerges clearly as the second factor. The older areas, positive loadings, are characterized by a high mean age, a high proportion over 65 years old, moved prior to 1960 and live in units built before 1940. In sharp contrast are the young areas, which moved after 1960 or within the past five years, as well as having a lower mean age. Map 2 testifies to the region's growth experience of the sixties. The rapid expansion of the southern portion of the region is quite evident with a strong concentration of "young" tracts. The outward expansion of Contra Costa and Alameda and a portion of Marin counties is also plainly present. The "old" portions of the region are much of San Francisco, the northern portion of Berkeley, and east Oakland. The retirement center in Walnut Creek (tract 351100) shows up as a suburban anomaly. The centers of the well-established towns like Santa Rosa, Napa, Yountville, and Vallejo are also "old."

The third factor is a poverty dimension, independent and in addition to the status factor. The presence of this factor emphasizes the power of the factorial ecology approach. Status is not simply high income. A variety of variables interacted to produce status as it is perceived in our society. Similarly poverty is not only the lack of money, nor the inverse of status, but a separate blending of a variable set. The poverty areas are indicated by a high proportion of the population being Black, with incomes less than 125% of the poverty level, and primarily employed by the service sector. The other end of the spectrum is a high proportion White or those of mixed parentage.

The fourth factor describes female participation in the labor force. It is characterized by a high percentage of females over 14 years old in the labor force, high employee to population ratio, and high clerical and sales workers. At first Map 4 shows no striking spatial pattern. However closer examination of factor four reveals a distinct locational pattern. High scoring tracts are located with major medical institutions and universities. The most significant concentration of high scoring tracts is the northern portion of San Francisco; and easy commute to downtown employment. Areas of significant male dominated employment display strong negative values. The Presidio, wharf, and south of Market areas are examples.

The fifth factor is dominated by the Hispanic population. Other high loading variables include percent of the dwelling units with more than one person per room (crowdedness), percent families earning less than \$5,000 a year, percent population under five years old, and percent families earning less than 125% of the poverty level. The other extreme (negative loadings) is composed of the percentage with high school education and a high mean school years attained. Map 5 clearly shows the well recognized Hispanic parts of the region: San Jose, the Mission District of San Francisco, Union City, Hayward and parts of Oakland.

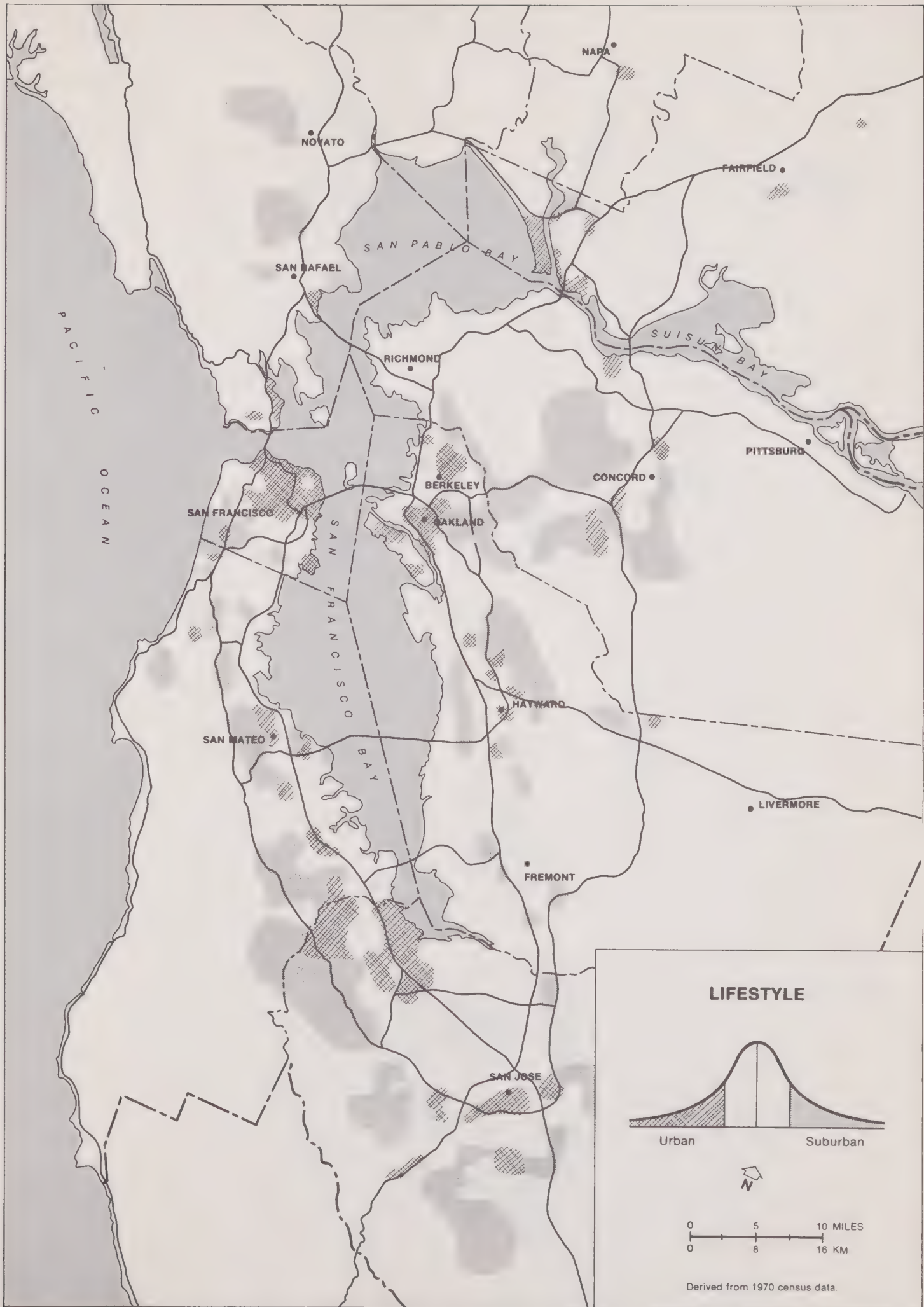
Chinese and other races form a distinct factor. The only high negative loading is percent of native born. The "Chinatowns" of San Francisco, Oakland and San Jose are clearly displayed on Map 6. The significant other races, such as the Filipino, emerge in the Western addition of San Francisco.

The seventh and final factor is characterized by a high persons per household ratio and a high percent in group quarters. Map 7 draws out many institutions, such as college dormitories, persons, and military barracks.

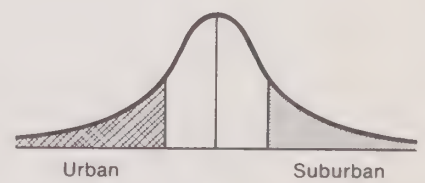
The seven factor solution fulfills the goals of social area analysis and provides the key descriptions of the social fabric:

- I Lifestyle - a choice of suburban or urban existence.
- II Stage-in-Family-Life Cycle - how people locate in social and physical space to meet their housing demands through the changes life brings.
- III Black Subpopulation - also describes families in economic distress and working in the low paying service sector.
- IV Women in the Labor Force - also indicates a high employee to resident ratio.
- V Hispanic Heritage subpopulation - also describes inadequate housing, poor education and many children.
- VI Asian subpopulation - dominated by Chinese, but other minority groups, such as Filipino, are described.
- VII Group Quarters - indicates special institutional housing arrangements.

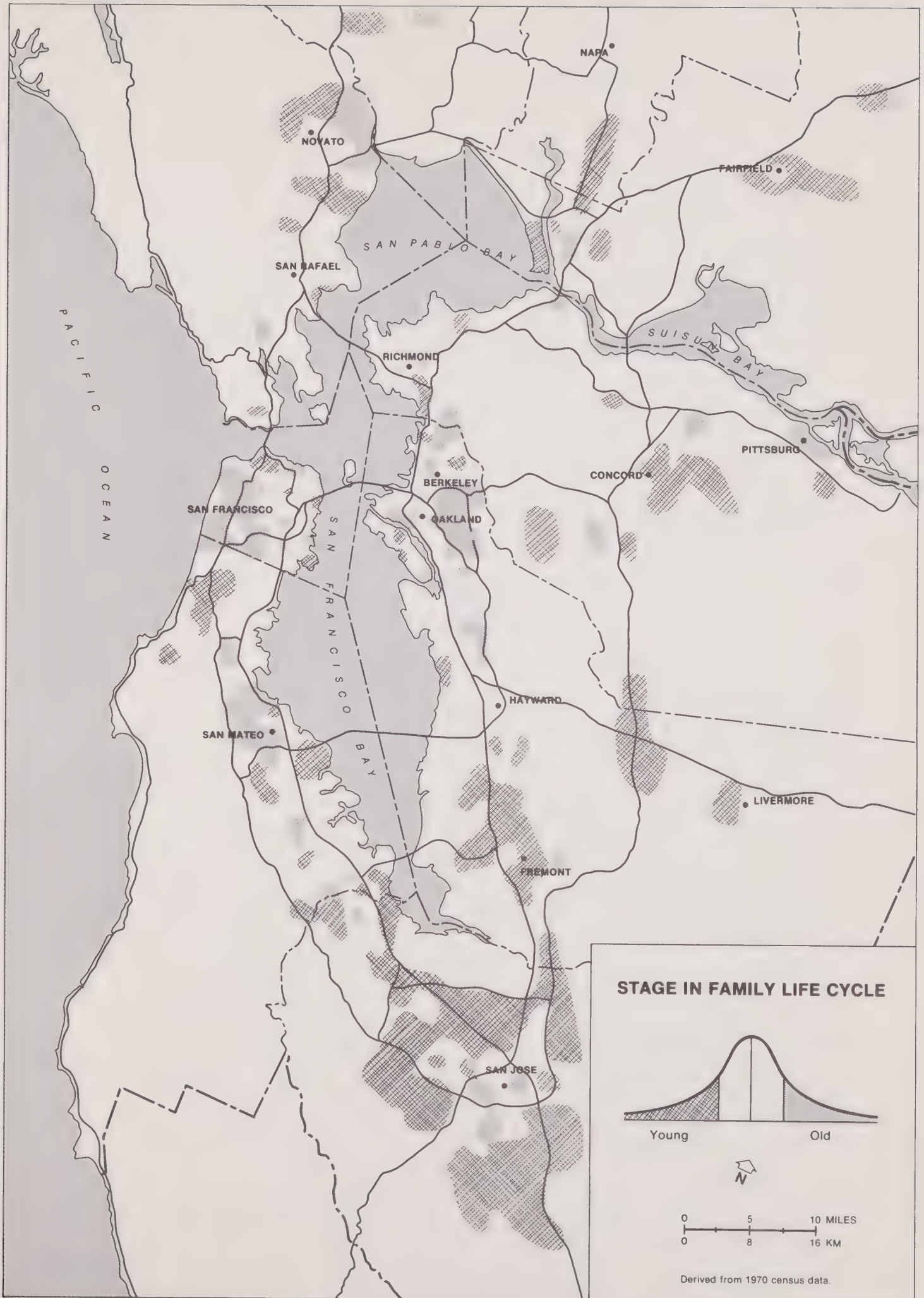
All of the factors are necessary to describe the social area.

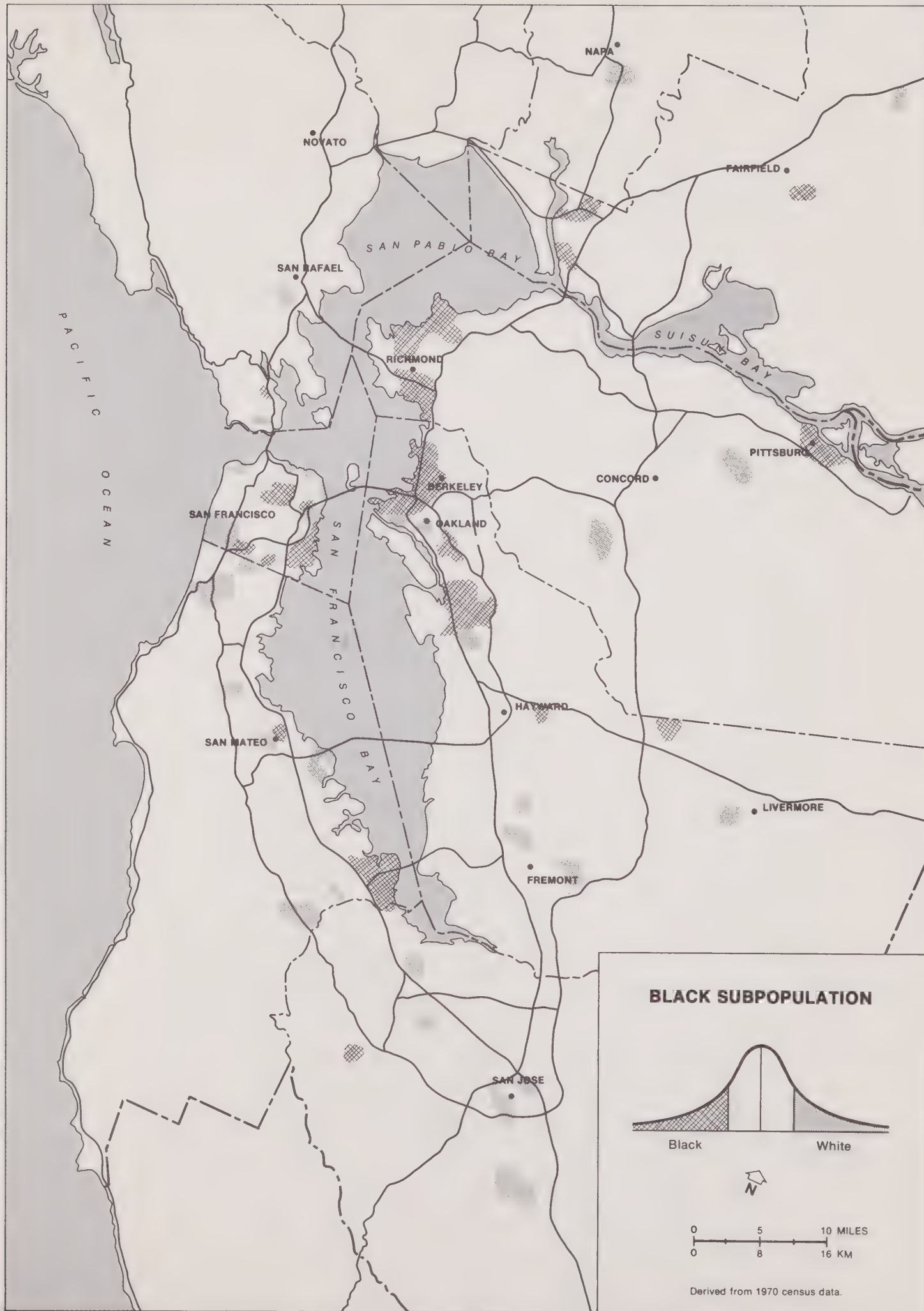


LIFESTYLE

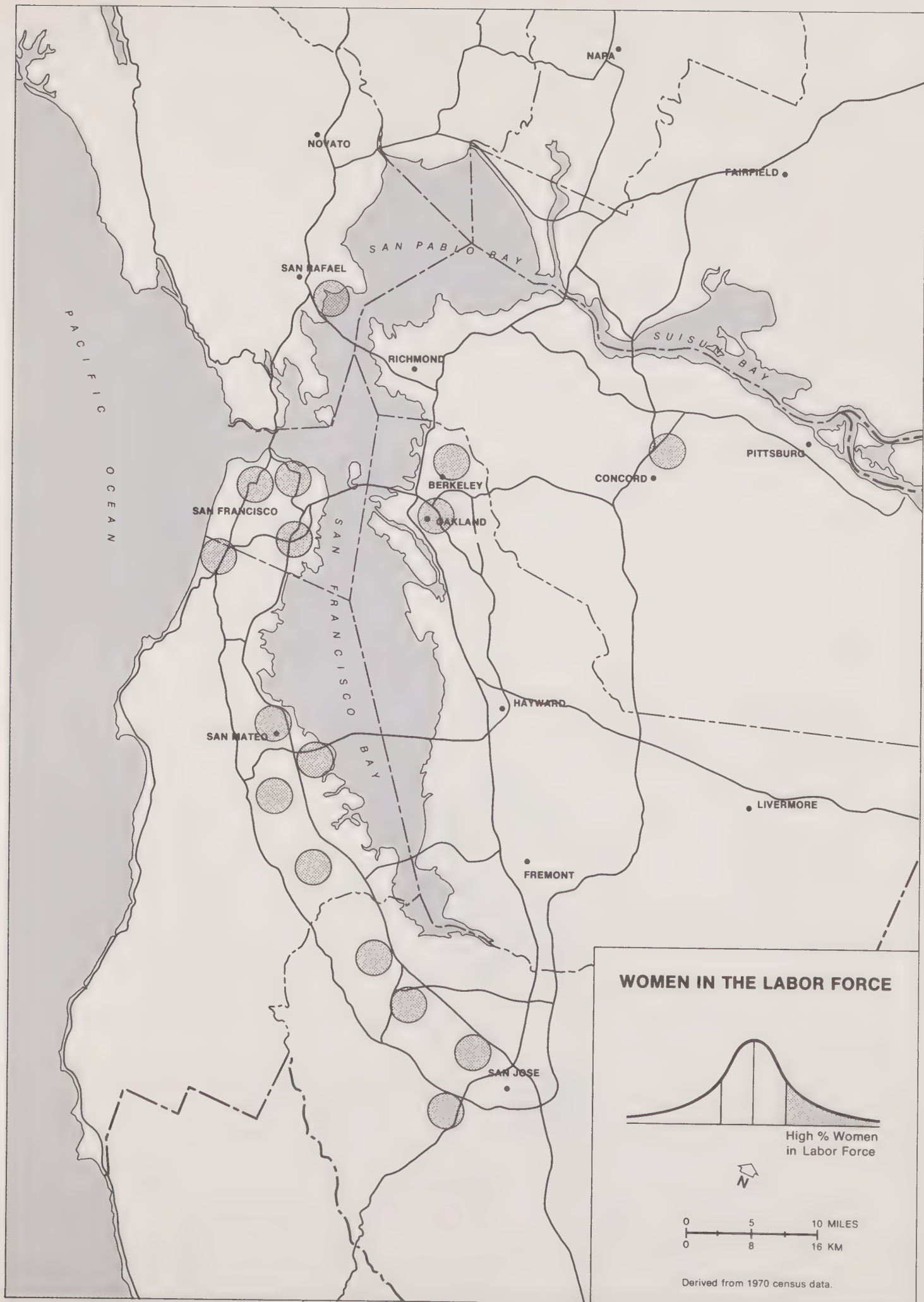


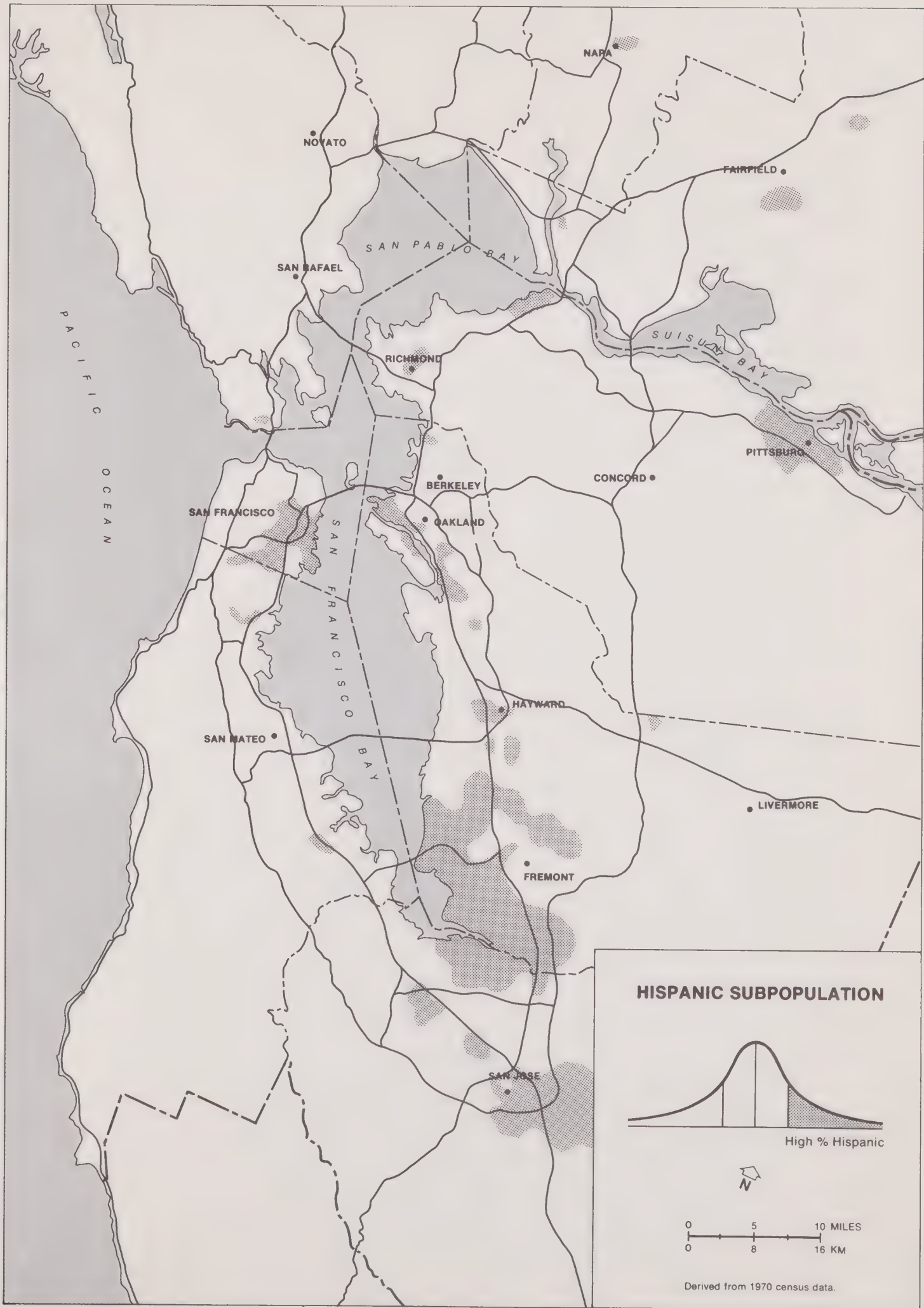
Derived from 1970 census data.

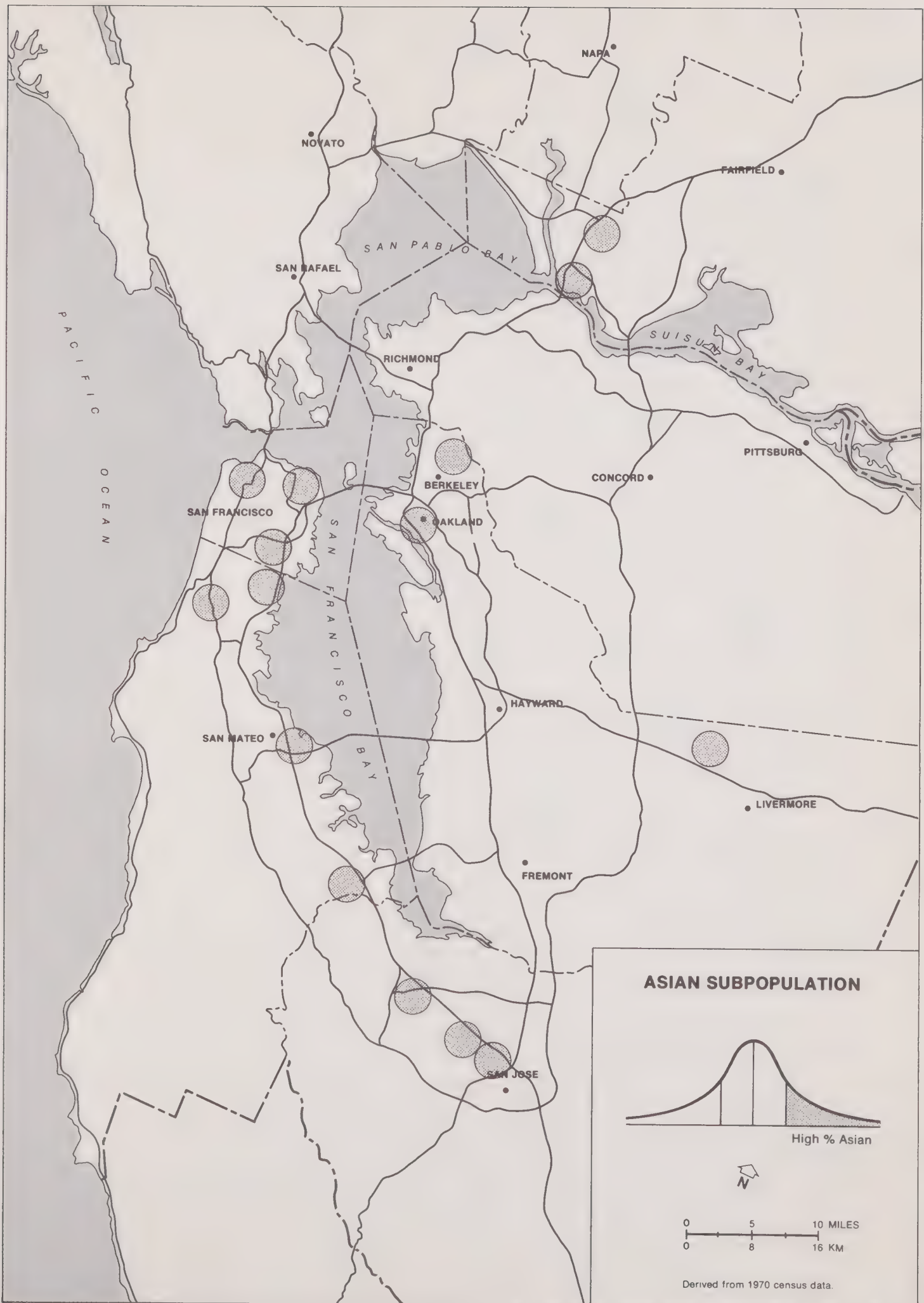


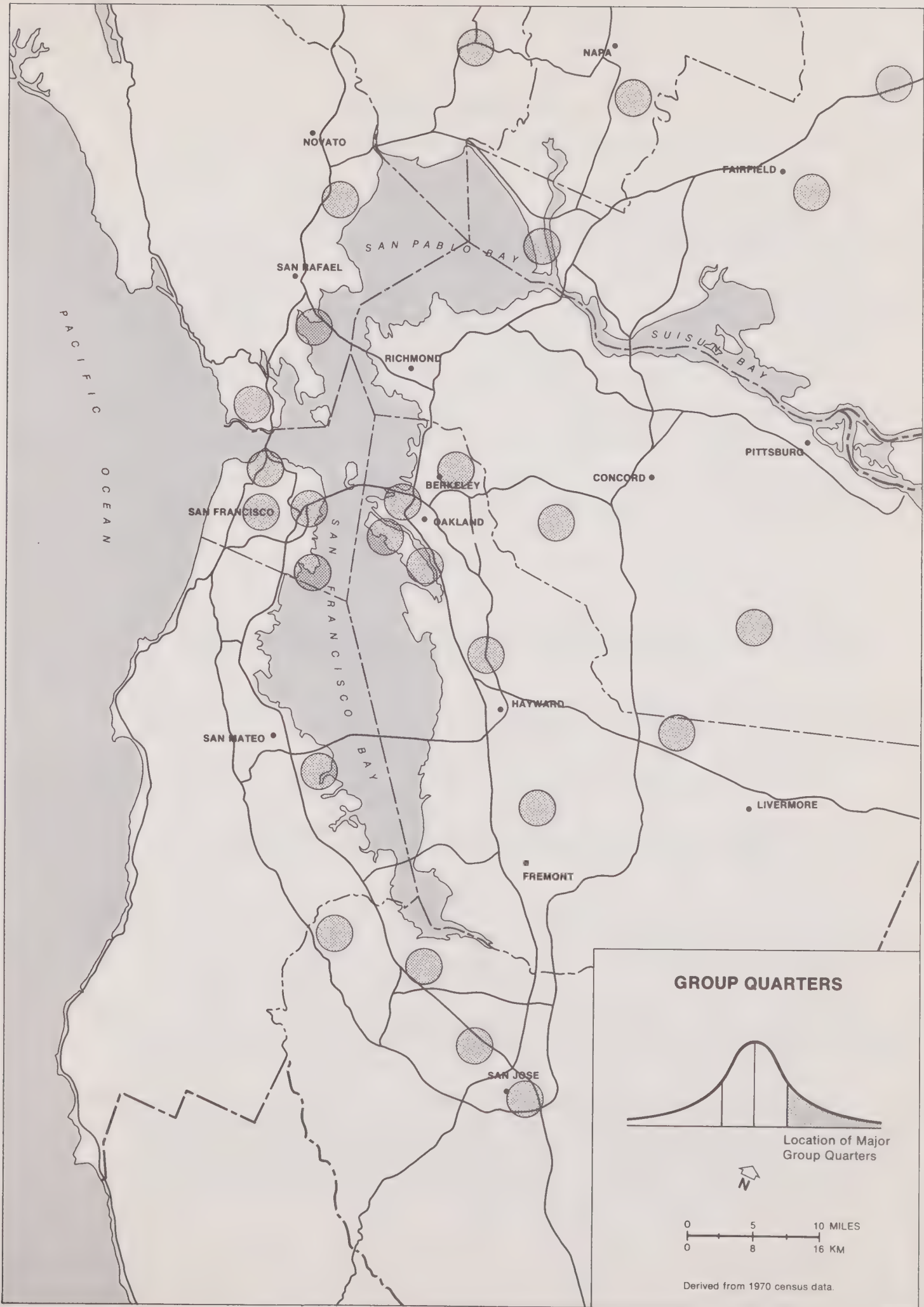


Derived from 1970 census data.









SUGGESTED BIBLIOGRAPHY

SOCIAL AREA ANALYSIS

- Adams, John, S., "Residential Structure of Midwestern Cities," Annals of the Association of American Geographers, LX (1970), pp. 37-62.
- Anderson, Theodore R. and Janice A. Egeland, "Spatial Aspects of Social Area Analysis," American Sociological Review, XXVI (1961), pp. 392-398.
- Bell, Wendell, "Economic, Family and Ethnic Status: An Empirical Test," American Sociological Review, XXII (1955), pp. 45-52.
- Berry, Brian J. L. and Philip H. Rees, "The Factorial Ecology of Calcutta," The American Journal of Sociology, LXXIV (1969), pp. 445-491.
- Burgess, Ernest W., "The Growth of the City," in R. E. Park, E. W. Burgess, and R. D. McKenzie, The City. Chicago: University of Chicago Press, 1925, pp. 47-62.
- Harman, Harry H., Modern Factor Analysis. Chicago: University of Chicago Press, 1960.
- Harris, Chauncy D. and Edward L. Ullman, "The Nature of Cities," The Annals of the American Academy of Political and Social Science, CCXLII (1945), pp. 7-17.
- Hartshorn, Truman A., "Inner City Residential Structure and Decline," Annals of the Association of American Geographers, LXI (1971), pp. 72-96.
- Hawley, Amos and Otis Dudley Duncan, "Social Area Analysis: A Critical Appraisal," Land Economics, XXXIII (1957), pp. 337-345.
- Hoyt, Homer, The Structure and Growth of Residential Neighborhoods in American Cities. Washington, D.C.: The Federal Housing Administration, 1939.
- Kain, John F., "The Journey-to-Work As a Determinant of Residential Location," in Alfred N. Page and Warren R. Seyfried (editors), Urban Analysis: Readings in Housing and Urban Development. Glenview, Illinois: Scott, Foresman and Company, 1970, pp. 207-226.
- Murdie, Robert, Factorial Ecology of Metropolitan Toronto, 1951-1961. Chicago: Research Paper Number 116, Department of Geography, University of Chicago, 1969.
- Rees, Philip H., "The Urban Envelope: Patterns and Dynamics of Population Density" and "Concepts of Social Space: Toward an Urban Social Geography," Chapters 9 and 10, respectively, of Brian J. L. Berry and Frank E. Horton, Geographic Perspectives on Urban Systems. Englewood Cliffs: Prentice Hall, 1970, pp. 276-394.

- Rummel, R. J., Applied Factor Analysis. Evanston, Illinois: Northwestern University Press, 1970.
- Schmid, Calvin F. and Kiyoshi Tagashira, "Ecological and Demographic Indices: A Methodological Analysis," Demography, I (1964), pp. 194-211.
- Simmons, James W., "Descriptive Models of Urban Land Use," Canadian Geographer, IX (1965), pp. 170-174.
- Sweetser, Frank L., Patterns of Change in the Social Ecology of Metropolitan Boston, 1950-1960. Boston: Division of Mental Hygiene, Massachusetts Department of Mental Health, 1962.
- Van Arsdol, Maurice, Santo F. Camilleria, and Calvin F. Schmid, "The Generality of Urban Social Area Indices," American Sociological Review, XXIII (1958), pp. 277-284.

4. COMMUNITY DEFINITION

The spatial delineation of "communities" is the critical step in the project. The areas must be suitable for demographic projections over a 15 to 20 year period, and appropriate to health planning interpretation and analysis. The use of social area analysis (described in section 2.) produces for each unit a measurement for each of the seven social indices: lifestyle, stage-in-life cycle, ethnic, and other factors. Developing the community definition is analogous to classifying tracts by their social characteristics. The scores are grouped to produce homogeneous subareas or communities. The resulting definition is homogeneous with respect to status, life cycle stage and ethnic composition and as such is an excellent basis for the project's analysis. The methodology is fully discussed below.

Tract Similarity Measurement

Grouping census tracts using some measurement of pair-wise association is a common, well established methodology in geography. The most common example of a measurement of association is simple euclidean distance, defined as:

$$d_{ij} = \left[\sum_r (s_{ir} - s_{jr})^2 \right]^{\frac{1}{2}} \quad (46)$$

for distance in r - space. The measurement is easy to interpret. If each tract is described in " r " ways, very similar tracts have a very small "distance" between them. As a pair of tracts are more dissimilar the distance increases. Since the dimensions resulting from the factor analysis are orthogonal, the factor scores for each tract can be used directly in computing the distance measurement.

Single Linkage Algorithms

The "single-linkage" algorithms are by far the most important of the grouping methods. Pairs of points are joined by steps according to a rule until all points are in one group. The method is conceptually simple and easy to develop. Since only single links are possible, clusters cannot overlap.

The most common single linkage methods in use are the centroid and Ward's grouping algorithms. Both begin with the generation of a n by n matrix of association between all n points. Initially each point or OTU (operational taxonomic unit) has $n-1$ "neighbors", decreasing with each step as groups are formed.

The centroid method joins at each step that pair of points (i,j) for which the measurement of association is a minimum. The appropriate i -th row and column and j -th row and column are deleted from the matrix and replaced by a new point which is the centroid of the pair (or group). The process is

repeated until all the points are grouped into one set. This method would be expected to work well when the data are already well patterned, but when groups are highly dispersed the grouping may become unstable. This is due to what is termed the "chaining" problem. Suppose there are several points in a row, evenly spaced. If one point is introduced, closer to one point than the others are to each other, as with point eight, the conditions are set up for "chaining." Points four and eight would be joined first, for

1 2 3 4 8 5 6 7

they are most similar. However, their centroid lies between four and eight and closer to five than to three. Point five would then be joined to the group (4,8) and a new centroid placed somewhere between (4,8) and five. Point six would next be joined and so on down the line. Only after seven was joined to the group would it be possible to join points one, two, and three. This process of the centroid dragging a group in one direction is the "chaining" problem. On a more general scale it could force a group to be dragged to a larger group when actually the former was more similar to a very different group. The centroid grouping works better on well-patterned or densely-packed data, but the chaining problem exists when data are dispersed.

In an attempt to develop a better algorithm, Ward developed a routine that examines the entire matrix and joins that pair which makes the minimum increment to the pooled within-group sum of the squared distances. Within-group distances are checked at each step. The matrix is updated as above, but uses the group mean. By minimizing the sum of the squared deviations about the group mean, the procedure can: (1) maintain groups of nearly equal size, (2) maintain groups of high density, and (3) develop groups that are spherical in shape. The algorithm is better than the centroid grouping mathematically, for it can work in less patterned data and draw out a clean grouping. But since it employs the group mean for updating groups it suffers to some degree from the same chaining problem as the simpler centroid method.

Operationally, the centroid and Ward's algorithms have several problems in common. Because of the larger size $(n \text{ by } n-1)/2$ of the distance matrix and the size of present computer storage, the sample size that can be handled operationally is very limited. Scanning the matrix at each $(n-1)$ step employs a great amount of computer time. Earlier programs could not handle over 220 points at a reasonable cost.

Neighborhood Limited Approach

A major design change is needed to handle a classification (grouping) of the 1,042 census tracts used in the factor analysis. If an algorithm stores only each point's topological neighbor, the very large distance matrices are not required. For example, for 440 zones, 96,580 storage bytes are needed. For 1,071 tracts 572,985 cells are required! If an algorithm is neighborhood limited the computer need only scan neighbor sets at each step of the grouping, minimizing computer time and storage. If only the ten "nearest" neighbors are needed, storage of the distance matrix for 1,042 tracts is reduced to 10,420 cells.

Neighbor Definition

A technical problem is neighbor definition. In geography the data set usually defines neighbor as spatially contiguous units. Contiguity is a relationship between two spatial partitions and not a characteristic of them. The relationship can generally be described in two manners. If a common boundary or line segment is chosen as the basis of determining contiguity, few neighbors develop. This shall be termed the Rook's case after the chess relationship. In Figure 13, area X has only the neighbors B, D, E, and G. The Queen's case is much less restrictive, and needs only one vertex in common to define contiguity. Point X then has the neighbor set A, B, C, D, E, F, G and H.

A new computer program which combines the virtues of the Ward's algorithm with the neighborhood limited concept was developed for the project. The description of GROUPT is in Part III, software documentation. To define communities the Queen's case is used below. Bridges are not considered as constituting legitimate contiguous points to define spatial neighbors.

Contiguity Matrix

The standard method of developing a contiguity matrix is simple. Tape a census tract outline map to a table. Number the tracts sequentially (1 to 1,042 in the analysis here). For each unit of analysis (census tract) record the sequence number of each tract that touches it, even if only at corners. The table must be keypunched as described in the GROUPT documentation. A short cut approach would be to record only four neighboring tracts. Table 30 shows that for the region, four neighbors is sufficient for the grouping analysis.

The Grouping Analysis

Normally the analyst would proceed directly to the GROUPT program to develop the social areas. However, the ABAG land use modeling system is tied to a traffic zone system based on aggregation of census tracts. The zonal definition was developed by the Metropolitan Transportation Commission (MTC). The region contains 440 zones. To make the various models compatible the tract information was converted to zones. For each zone an average score was calculated for each of the seven dimensions based on the tracts within that zone. The reduction of the 1,042 tracts to 440 zones is only for ABAG's convenience. The examples below deal with zones, but the method steps apply equally to the normal tract level processing.

The GROUPT program was run with the zonal factor score means for all seven factors, and the complete zonal contiguity matrix for up to nine neighbors, for all 440 zones. The program produced the grouping, beginning with 440 "groups" of one zone each, and joining groups until step 422 when the contiguity matrix prevented going further. At each step the algorithm printed the zones joined, the cumulative step join, and the step join. The results are Table 31. As displayed in this table, at first the algorithm joins extremely similar zones; the "distance" between them is zero. The distance of the join increases slowly, since similar zones are

FIGURE 13
ROOKS CASE

A	B	C
D	X	E
F	G	H

TABLE 30
NEIGHBOR DISTRIBUTION

# neighbors	# tracts	% of all tracts
0	0	0
1	170	16.3
2	297	28.5
3	280	26.8
4	154	14.8
5	76	7.3
6	19	1.8
7	4	.4
8	4	.4
9	1	.1
10	1	.1

TABLE 31
GROUPING OF 440 ZONES

SECOND PASS AT GROUPING ALL 440 ZONES IN SEVEN SPACE

NUMBER OF TRACTS TO BE GROUPED	440
NUMBER OF VARIABLES	7
MAXIMUM NUMBER OF CONTIGUOUS TRACTS	9
CONTIGUITY LISTING	0
FORMAT OF SCORE INPUT: (3X,7F7.4)	

CLUSTERING CRITERIA - WARDS ERROR-SUM-OF-SQUARES OBJECTIVE FUNCTION

CYCLE TRACT 1 TRACT 2 CRITERIA VALUE DIFFERENCE
NUMBER

(1)	(2)	(3)	(4)	(5)
1	424	426	-0.0	-0.0
2	425	427	-0.0	0.0
3	7	8	0.1	0.1
4	95	105	0.1	0.0
5	407	410	0.1	0.0
6	263	251	0.1	0.0
7	101	103	0.1	0.0
8	243	283	0.1	0.0
9	23	26	0.1	0.0
10	89	86	0.1	0.0
11	284	244	0.1	0.0
12	149	151	0.1	0.0
13	311	315	0.1	0.0
14	330	332	0.1	0.0
15	195	203	0.1	0.0
16	298	300	0.1	0.0
17	276	248	0.1	0.0
18	187	209	0.1	0.0

TABLE 31 (Continued)

19	154	182	0.1	0.0
20	204	210	0.1	0.0
21	375	374	0.2	0.0
22	351	354	0.2	0.0
23	356	357	0.2	0.0
24	378	403	0.2	0.0
25	180	183	0.2	0.0
26	152	158	0.2	0.0
27	226	233	0.2	0.0
28	50	64	0.2	0.0
29	83	82	0.2	0.0
30	406	409	0.2	0.0
31	120	124	0.2	0.0
32	394	390	0.2	0.0
33	196	197	0.2	0.0
34	360	359	0.2	0.0
35	331	355	0.2	0.0
36	21	12	0.2	0.0
37	164	167	0.2	0.0
38	35	39	0.2	0.0
39	372	371	0.2	0.0
40	243	246	0.2	0.0
41	326	339	0.2	0.0
42	266	269	0.2	0.0
43	287	289	0.2	0.0
44	176	179	0.2	0.0
45	234	236	0.2	0.0
46	223	238	0.2	0.0
47	189	192	0.3	0.0
48	159	173	0.3	0.0
49	94	99	0.3	0.0
50	44	43	0.3	0.0
51	30	31	0.3	0.0
52	186	205	0.3	0.0
53	347	362	0.3	0.0
54	175	190	0.3	0.0
55	5	7	0.3	0.0
56	282	290	0.3	0.0
57	221	227	0.3	0.0
58	440	439	0.3	0.0
59	107	97	0.3	0.0
60	317	319	0.3	0.0
61	215	213	0.3	0.0
62	122	121	0.3	0.0
63	123	150	0.4	0.0
64	109	114	0.4	0.0
65	157	153	0.4	0.0
66	250	249	0.4	0.0
67	32	42	0.4	0.0
68	144	140	0.4	0.0
69	348	361	0.4	0.0
70	318	334	0.4	0.0

TABLE 31 (Continued)

71	384	397	0.4	0.0
72	436	437	0.4	0.0
73	386	389	0.4	0.0
74	202	201	0.4	0.0
75	323	314	0.4	0.0
76	55	54	0.4	0.0
77	170	178	0.4	0.0
78	217	219	0.5	0.0
79	96	106	0.5	0.0
80	163	166	0.5	0.0
81	191	193	0.5	0.0
82	296	295	0.5	0.0
83	45	46	0.5	0.0
84	92	131	0.5	0.0
85	417	423	0.5	0.0
86	247	241	0.5	0.0
87	18	27	0.5	0.0
88	212	214	0.5	0.0
89	243	245	0.5	0.0
90	127	138	0.6	0.0
91	101	100	0.6	0.0
92	176	184	0.6	0.0
93	321	324	0.6	0.0
94	125	135	0.6	0.0
95	276	278	0.6	0.0
96	128	133	0.6	0.0
97	232	235	0.7	0.0
98	376	399	0.7	0.0
99	20	15	0.7	0.0
100	256	259	0.7	0.0
101	240	247	0.7	0.0
102	341	340	0.7	0.0
103	333	336	0.7	0.0
104	305	285	0.7	0.0
105	104	102	0.7	0.0
106	194	199	0.7	0.0
107	252	265	0.8	0.0
108	19	18	0.8	0.0
109	307	325	0.8	0.0
110	44	24	0.8	0.0
111	280	279	0.9	0.0
112	118	117	0.9	0.0
113	414	413	0.9	0.0
114	365	364	0.9	0.0
115	195	196	0.9	0.0
116	113	119	0.9	0.0
117	304	303	0.9	0.0
118	160	161	0.9	0.0
119	270	267	0.9	0.0
120	155	174	0.9	0.0
121	91	92	1.0	0.0
122	404	401	1.0	0.0
123	387	388	1.0	0.0

TABLE 31 (Continued)

124	65	69	1.0	0.0
125	74	78	1.0	0.0
126	326	318	1.1	0.0
127	255	258	1.1	0.0
128	232	234	1.1	0.0
129	93	96	1.1	0.0
130	348	347	1.2	0.0
131	273	275	1.2	0.0
132	80	216	1.2	0.0
133	35	37	1.2	0.0
134	438	419	1.2	0.0
135	9	20	1.3	0.0
136	108	75	1.3	0.0
137	204	200	1.4	0.1
138	240	242	1.4	0.0
139	148	123	1.4	0.0
140	89	87	1.4	0.0
141	320	323	1.4	0.0
142	191	198	1.4	0.0
143	367	379	1.4	0.0
144	165	168	1.4	0.0
145	47	52	1.5	0.0
146	33	32	1.5	0.0
147	276	277	1.5	0.0
148	345	352	1.5	0.0
149	224	239	1.6	0.0
150	305	316	1.6	0.0
151	59	62	1.6	0.0
152	145	147	1.6	0.0
153	58	57	1.7	0.0
154	343	338	1.7	0.0
155	329	344	1.7	0.0
156	176	181	1.7	0.0
157	218	220	1.8	0.1
158	89	88	1.8	0.0
159	435	434	1.8	0.0
160	95	107	1.8	0.0
161	237	223	1.8	0.0
162	284	243	1.9	0.0
163	264	271	1.9	0.0
164	382	428	1.9	0.0
165	375	377	1.9	0.0
166	270	229	1.9	0.0
167	33	41	2.0	0.0
168	21	11	2.1	0.1
169	19	28	2.1	0.0
170	45	44	2.1	0.1
171	73	76	2.2	0.0
172	302	309	2.2	0.0
173	101	93	2.2	0.0
174	127	134	2.2	0.0
175	149	148	2.2	0.0
176	50	71	2.3	0.0

TABLE 31 (Continued)

177	175	177	2.3	0.0
178	195	202	2.3	0.0
179	387	386	2.3	0.0
180	53	59	2.3	0.0
181	73	77	2.4	0.0
182	368	380	2.4	0.0
183	186	85	2.4	0.0
184	348	363	2.4	0.0
185	308	310	2.4	0.0
186	109	110	2.4	0.0
187	60	63	2.4	0.0
188	282	281	2.4	0.0
189	9	22	2.5	0.0
190	154	188	2.5	0.0
191	35	38	2.5	0.0
192	276	250	2.5	0.0
193	335	326	2.5	0.0
194	163	164	2.5	0.0
195	204	293	2.6	0.0
196	320	322	2.6	0.0
197	25	30	2.6	0.1
198	266	257	2.7	0.1
199	396	414	2.7	0.0
200	297	204	2.7	0.0
201	262	264	2.8	0.1
202	5	21	2.8	0.0
203	351	350	2.9	0.1
204	286	288	2.9	0.0
205	270	228	3.0	0.0
206	231	230	3.0	0.0
207	385	395	3.0	0.0
208	317	305	3.0	0.0
209	129	137	3.1	0.0
210	189	191	3.1	0.0
211	406	412	3.1	0.0
212	411	440	3.1	0.0
213	90	94	3.1	0.0
214	74	79	3.1	0.0
215	176	170	3.3	0.2
216	118	116	3.4	0.1
217	343	337	3.5	0.0
218	287	299	3.5	0.0
219	394	391	3.5	0.0
220	66	68	3.5	0.0
221	127	139	3.6	0.0
222	297	294	3.6	0.1
223	400	378	3.7	0.1
224	360	358	3.8	0.1
225	255	292	3.9	0.2
226	329	345	4.0	0.0
227	104	109	4.0	0.1
228	276	280	4.0	0.0
229	262	260	4.0	0.0

TABLE 31 (Continued)

230	232	224	4.1	0.0
231	263	253	4.1	0.0
232	17	29	4.2	0.1
233	365	366	4.2	0.0
234	405	407	4.2	0.0
235	416	438	4.2	0.0
236	120	122	4.3	0.1
237	35	48	4.3	0.0
238	351	353	4.4	0.1
239	154	156	4.4	0.0
240	372	369	4.4	0.0
241	9	16	4.4	0.0
242	157	152	4.5	0.0
243	141	160	4.5	0.1
244	91	95	4.6	0.1
245	367	373	4.6	0.0
246	186	89	4.6	0.0
247	33	34	4.6	0.0
248	5	13	4.7	0.1
249	400	402	4.7	0.0
250	297	226	4.7	0.0
251	70	50	4.9	0.2
252	83	81	5.0	0.0
253	187	206	5.0	0.1
254	47	49	5.4	0.3
255	112	126	5.4	0.0
256	58	53	5.4	0.0
257	394	392	5.5	0.0
258	273	274	5.5	0.0
259	330	321	5.5	0.0
260	104	111	5.6	0.1
261	127	136	5.7	0.1
262	129	132	5.7	0.0
263	176	185	5.7	0.0
264	317	240	5.7	0.0
265	215	212	5.8	0.1
266	73	108	5.8	0.0
267	252	254	5.8	0.0
268	1	45	6.0	0.1
269	266	270	6.0	0.1
270	163	165	6.1	0.0
271	154	155	6.1	0.0
272	218	222	6.1	0.0
273	35	36	6.2	0.1
274	287	301	6.2	0.0
275	311	313	6.2	0.0
276	189	194	6.5	0.3
277	23	25	6.7	0.2
278	381	383	6.7	0.0
279	348	360	6.7	0.0
280	172	159	6.8	0.0
281	262	272	6.9	0.2
282	74	80	7.0	0.0
283	335	333	7.1	0.1

TABLE 31 (Continued)

284	416	418	7.2	0.1
285	393	384	7.2	0.0
286	195	297	7.6	0.4
287	2	5	7.9	0.3
288	232	237	8.0	0.2
289	118	115	8.1	0.0
290	66	72	8.1	0.0
291	113	120	8.2	0.1
292	417	422	8.2	0.0
293	276	282	8.3	0.1
294	35	40	8.8	0.4
295	47	51	8.8	0.0
296	125	128	8.8	0.0
297	317	328	8.8	0.0
298	90	101	8.9	0.1
299	261	252	9.1	0.2
300	405	406	9.3	0.2
301	157	175	9.6	0.3
302	141	146	9.8	0.2
303	367	375	9.9	0.2
304	330	329	10.0	0.0
305	435	430	10.0	0.1
306	255	256	10.1	0.1
307	385	387	10.5	0.4
308	304	302	10.6	0.1
309	286	287	10.9	0.2
310	376	398	10.9	0.0
311	308	312	11.0	0.2
312	2	14	11.1	0.1
313	149	91	11.2	0.1
314	320	311	11.5	0.3
315	70	65	11.7	0.2
316	33	58	12.0	0.3
317	73	83	12.2	0.2
318	263	273	12.3	0.1
319	335	343	12.3	0.0
320	9	19	12.8	0.4
321	231	261	13.0	0.3
322	144	143	13.2	0.1
323	195	268	13.2	0.0
324	348	365	13.3	0.1
325	351	331	13.7	0.4
326	411	416	14.5	0.8
327	393	429	14.5	0.1
328	125	145	14.6	0.1
329	317	284	15.0	0.3

TABLE 31 (Continued)

330	163	176	15.0	0.1
331	186	90	15.3	0.3
332	420	435	15.7	0.4
333	307	342	15.8	0.1
334	396	424	15.9	0.1
335	221	266	16.4	0.4
336	436	433	16.4	0.0
337	1	23	16.6	0.2
338	169	154	17.2	0.6
339	118	127	17.8	0.5
340	405	415	18.5	0.8
341	376	400	18.6	0.1
342	215	74	18.8	0.2
343	368	370	19.2	0.4
344	104	113	19.2	0.0
345	66	67	19.2	0.0
346	33	60	19.4	0.1
347	367	394	20.0	0.7
348	157	180	20.4	0.4
349	129	130	20.5	0.1
350	35	47	20.9	0.4
351	172	141	21.2	0.3
352	187	195	21.4	0.3
353	70	55	21.5	0.1
354	149	98	22.2	0.6
355	335	341	22.2	0.0
356	262	255	22.2	0.0
357	218	232	22.6	0.4
358	9	17	24.4	1.7
359	286	276	24.9	0.5
360	330	351	25.6	0.8
361	304	298	26.6	1.0
362	144	142	26.8	0.1
363	163	189	28.4	1.7
364	348	356	29.6	1.1
365	417	411	29.6	0.1
366	308	320	30.0	0.4
367	215	296	30.5	0.5
368	263	231	31.0	0.5
369	118	112	32.1	1.1
370	385	393	33.9	1.8
371	372	367	36.6	2.7
372	6	2	38.5	1.9
373	376	404	38.5	0.0
374	1	33	38.5	0.0
375	73	104	41.7	3.1
376	125	129	42.3	0.7
377	396	382	43.6	1.2
378	405	408	43.6	0.1
379	307	335	43.9	0.3
380	218	317	43.9	0.0
381	169	172	45.2	1.3
382	157	186	45.7	0.5
383	304	306	46.2	0.5

TABLE 31 (Continued)

384	66	70	46.4	0.2
385	286	291	47.9	1.5
386	420	432	48.1	0.2
387	217	221	50.1	2.0
388	417	436	55.3	5.2
389	263	262	59.5	4.3
390	308	330	60.9	1.4
391	163	187	65.4	4.4
392	6	9	67.0	1.6
393	1	35	67.4	0.5
394	149	157	73.7	6.3
395	73	215	75.6	1.9
396	385	368	79.7	4.1
397	144	125	80.3	0.6
398	372	376	80.9	0.6
399	118	169	83.5	2.5
400	304	218	100.4	17.0
401	308	348	101.7	1.3
402	381	396	101.8	0.1
403	417	431	107.2	5.5
404	1	66	119.4	12.2
405	286	263	132.8	13.4
406	372	405	149.2	16.4
407	307	308	150.4	1.2
408	73	163	153.7	3.3
409	118	144	198.8	45.1
410	381	385	205.0	6.2
411	1	6	208.3	3.2
412	217	304	217.8	9.6
413	149	73	261.1	43.3
414	417	420	275.3	14.2
415	286	307	304.2	28.8
416	118	372	401.0	96.8
417	207	149	456.1	55.1
418	217	286	547.8	91.7
419	381	417	550.3	2.6
420	118	207	963.2	412.9
421	118	217	1534.9	571.6
422	381	118	2271.2	736.4

No further grouping of tracts was possible under contiguity used.

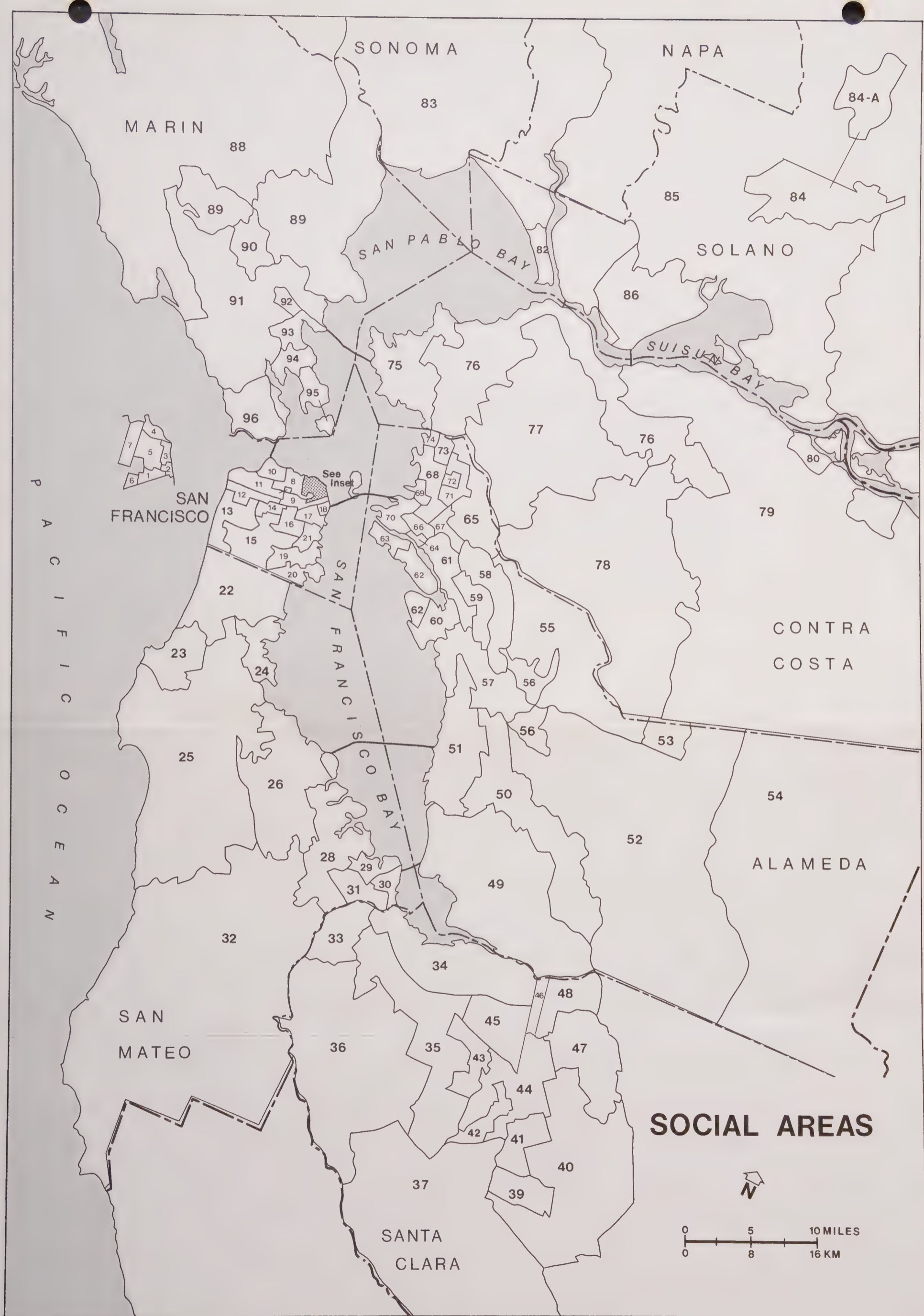
Explanation of output:

Column 1 - cycle number, from 1 to N-1,

Column 2 and 3 - sequence code of tract that was most similar to tract given in Column 3,

Column 4 - present value of clustering criteria,

Column 5 - change in value since previous cycle.



grouped into larger groups. If the step join suddenly increases, and then continues at a lower level for a few steps, it indicates that two slightly dissimilar groups were joined. The analyst should examine the groupings at the step previous to the larger join, for it is one possible optimal combination of strongly identified homogeneous groups and isolates.

A special program, GROUPEM, was written to allow the analyst to study the complete grouping. Because of the large number of zones involved, it would be expensive and useless to write out the group memberships and isolates at each step. GROUPEM examines the grouping at a given step, writes out group memberships and means on each dimension, and lists the isolates.

The first jump is at step 358 where the step join exceeds one (greater than one standard deviation for these normalized, standardized scores). Other sharp changes are at steps 371, 375, 394, 400, and 409. The groupings were analyzed with GROUPEM at the steps prior to these jumps. The groupings after step 358 tended to produce several large groups and a series of persistent isolates. Step 357, selected for the final grouping, produced 96 homogeneous social areas. Map 8 is the result and is the basis of the small area population projections needed for health planning purposes.

To facilitate understanding the social areas a profile book was prepared. It contains for each area:

1. the social area mean on each of the seven dimensions.
2. the census tract membership list.
3. for each member tract a listing of the actual census data for each variable that contributed significantly to each of the seven dimensions. The data listing is arranged by factor.

Table 32 is an example. When a score value is less than -1.00 or greater than +1.00, the area is significantly different from more than two-thirds of the other areas. It is statistically different from the regional average. Note in Table 32 that area 58 has large magnitude values on factors I, II, and VII. A high positive value for the first dimension means that the area is very suburban. The actual indicator values given for each member tract display the expected features. Tract 407800 in Oakland, for example, has 79 percent of the housing in single unit structures, 66 percent owner occupied, and 62 percent commute by car. The mean home value is a high \$23,573 (in 1969 dollars). By the value for the second factor, social area 58 is elderly. Scanning the raw indicators for census tracts 404800, 406600 and 406700 note that the percent over 65 years old is 20, 16 and 17 respectively. The high value for the group quarters dimension, factor VII is primarily due to tract 432900 which is 97 percent group quarters. It contains a juvenile home. The area is near the regional averages for the remaining dimensions.

TABLE 32

SOCIAL AREA # 58

	I	II	III	IV	V	VI	VII
SCORES:	1.516	-1.552	-0.498	-0.579	-0.394	-0.242	4.194
PROFILE							

SOCIAL AREA PROFILES BY CENSUS TRACT

CENSUS TRACT #404800.

OAKLAND

%HU 1 UNIT	57.580		
%OWN OCCPD	49.690		
HOME \$/INC	1.410		
%2+CARS	29.670		
AVG FAMILY S	1.870		
%INC>\$15000	20.140		
MEAN HOME \$	23132.000		
%POP <18 YRS	19.830		
%CMMT-AUTO	80.020		
%HU 3+ UNIT	38.120		
%HH FEMALE H	33.470		
%HU >1960	21.460		
%MOVED-1965	53.000		
%POP <5 YRS	6.650		
%MOVED<1960	34.720		
MEAN AGE	41.190		
%POP >65 YRS.	19.870		
%HU <1940	50.640		
%BLACK	0.580		
INC<125%POOR.	0.040		
%SERVICES	7.600		
%INC<\$5000	8.220		
%WHITE	93.930		
%OF MIXED	21.520		
%FEMALE>14LF.	37.240		
EMPLOY/POP	0.430		
%CLERIC+SALE.	33.370		
%CHICANO	5.930		
%DU>1.01/RM	1.760		
INC<125%POOR.	0.040		
%INC<\$5000	8.220		
%POP <5 YRS	6.650		
%>25 W HS	64.550		
MEAN SCHL YR.	11.700		
%CHINESE	2.840		
%OTHER RACES.	5.490		
%OF NATIVE	65.340		
PERSON/HH	2.290		
%GROUP Q	0.0		

CENSUS TRACT #406600.

OAKLAND

%HU 1 UNIT	47.460		
%OWN OCCPD	41.450		
HOME \$/INC	0.890		
%2+CARS	27.710		
AVG FAMILY S	2.150		
%INC>\$15000	15.070		
MEAN HOME \$	15902.398		
%POP <18 YRS	25.400		
%CMMT-AUTO	75.230		
%HU 3+ UNIT	40.930		
%HH FEMALE H	30.000		
%HU >1960	25.360		
%MOVED-1965	57.450		
%POP <5 YRS	9.140		
%MOVED<1960	25.230		
MEAN AGE	36.880		
%POP >65 YRS.	16.440		
%HU <1940	46.380		
%BLACK	4.700		
INC<125%POOR.	0.040		
%SERVICES	9.720		
%INC<\$5000	9.430		
%WHITE	85.450		
%OF MIXED	23.760		
%FEMALE>14LF.	49.170		
EMPLOY/POP	0.430		
%CLERIC+SALE.	37.380		
%CHICANO	10.360		
%DU>1.01/RM	4.210		
INC<125%POOR.	0.040		
%INC<\$5000	9.430		
%POP <5 YRS	9.140		
%>25 W HS	57.680		
MEAN SCHL YR.	11.100		
%CHINESE	4.080		
%OTHER RACES.	9.850		
%OF NATIVE	65.070		
PERSON/HH	2.500		
%GROUP Q	0.660		

CENSUS TRACT #406700.

OAKLAND

%HU 1 UNIT	85.340
%OWN OCCPD	71.010
HOME \$/INC	1.720
%2+CARS	39.010
AVG FAMILY S	2.230
%INC>\$15000	26.750
MEAN HOME \$	25253.602
%POP <18 YRS	21.870
%CMMT-AUTO	81.260
%HU 3+ UNIT	11.770
%HH FEMALE H	23.120
%HU >1960	13.220
%MOVED-1965	39.810
%POP <5 YRS	6.360
%MOVED<1960	37.670
MEAN AGE	40.660
%POP >65 YRS.	16.690

%HU <1940	53.010		
%BLACK	1.240		
INC<125%POOR.	0.040		
%SERVICES	7.260		
%INC<\$5000	7.900		
%WHITE	93.810		
%OF MIXED	27.070		
%FEMALE>14LF.	41.710		
EMPLOY/POP	0.450		
%CLERIC+SALE.	33.250		
%CHICANO	6.730		
%DU>1.01/RM	2.180		
INC<125%POOR.	0.040		
%INC<\$5000	7.900		
%POP <5 YRS	6.360		
%>25 W HS	65.490		
MEAN SCHL YR.	11.600		
%CHINESE	3.040		
%OTHER RACES.	4.950		
%OF NATIVE	62.250		
PERSON/HH	2.520		
%GROUP Q	0.950		

%HU 1 UNIT	79.340		
%OWN OCCPD	71.780		
HOME \$/INC	1.700		
%2+CARS	35.970		
AVG FAMILY S	2.260		
%INC>\$15000	27.310		
MEAN HOME \$	23138.898		
%POP <18 YRS	22.420		
%CMMT-AUTO	73.350		
%HU 3+ UNIT	15.890		
%HH FEMALE H	27.070		
%HU >1960	13.110	
%MOVED-1965	40.610	
%POP <5 YRS	6.210	
%MOVED<1960	37.780	
MEAN AGE	41.480	
%POP >65 YRS.	17.620	
%HU <1940	51.110	
%BLACK	0.420	
INC<125%POOR.	0.050	
%SERVICES	11.380	
%INC<\$5000	8.520	
%WHITE	93.310	
%OF MIXED	26.780	
%FEMALE>14LF.	38.620	
EMPLOY/POP	0.440	
%CLERIC+SALE.	24.520	
%CHICANO	12.800	
%DU>1.01/RM	1.110	
INC<125%POOR.	0.050	
%INC<\$5000	8.520	
%POP <5 YRS	6.210	
%>25 W HS	64.530	
MEAN SCHL YR.	11.100	
%CHINESE	3.530	
%OTHER RACES.	6.280	
%OF NATIVE	55.890	
PERSON/HH	2.550	
%GROUP Q	0.0	

OAKLAND

146

CENSUS TRACT #407000.

OAKLAND

%HU 1 UNIT	56.830		
%OWN OCCPD	42.130		
HOME \$/INC	1.050		
%2+CARS	21.480		
AVG FAMILY S	1.950		
%INC>\$15000	9.680		
MEAN HOME \$	15276.500		
%POP <18 YRS	21.670		
%CMMT-AUTO	72.880		
%HU 3+ UNIT	31.750		
%HH FEMALE H	33.470		
%HU >1960	17.200	
%MOVED-1965	51.240	
%POP <5 YRS	9.680	
%MOVED<1960	28.940	
MEAN AGE	39.230	
%POP >65 YRS	19.370	
%HU <1940	58.070	
%BLACK	6.020	
INC<125%POOR.	0.090	
%SERVICES	11.980	
%INC<\$5000	16.000	
%WHITE	85.530	
%OF MIXED	19.840	
%FEMALE>14LF.	37.160	
EMPLOY/POP	0.390	
%CLERIC+SALE.	34.520	
%CHICANO	18.590	
%DU>1.01/RM	4.810	
INC<125%POOR.	0.090	
%INC<\$5000	16.000	
%POP <5 YRS	9.680	
%>25 W HS	51.560	
MEAN SCHL YR.	10.300	
%CHINESE	3.110	
%OTHER RACES.	8.450	
%OF NATIVE	70.350	
PERSON/HH	2.400	
%GROUP Q	1.850	

CENSUS TRACT #407800.

OAKLAND

%HU 1 UNIT	79.300		
%OWN OCCPD	66.950		
HOME \$/INC	1.600		
%2+CARS	34.080		
AVG FAMILY S	2.550		
%INC>\$15000	24.380		
MEAN HOME \$	23573.699		
%POP <18 YRS	26.090		
%CMMT-AUTO	62.470		
%HU 3+ UNIT	18.410		
%HH FEMALE H	28.350		
%HU >1960	11.140	
%MOVED-1965	37.820	
%POP <5 YRS	8.230	
%MOVED<1960	31.650	
MEAN AGE	34.220	
%POP >65 YRS.	9.220	
%HU <1940	44.680	
%BLACK	12.510	
INC<125%POOR.	0.070	
%SERVICES	12.930	
%INC<\$5000	8.590	
%WHITE	82.300	
%OF MIXED	17.860	
%FEMALE>14LF.	45.030	
EMPLOY/POP	0.460	
%CLERIC+SALE.	34.450	
%CHICANO	6.810	
%DU>1.01/RM	3.760	
INC<125%POOR.	0.070	
%INC<\$5000	8.590	
%POP <5 YRS	8.230	
%>25 W HS	66.770	
MEAN SCHL YR.	11.600	
%CHINESE	4.360	
%OTHER RACES.	5.190	
%OF NATIVE	81.960	
PERSON/HH	3.210	
%GROUP Q	18.210	

CENSUS TRACT #407900.

OAKLAND

%HU 1 UNIT	89.090		
%OWN OCCPD	75.280		
HOME \$/INC	1.740		
%2+CARS	44.760		
AVG FAMILY S	2.430		
%INC>\$15000	28.240		
MEAN HOME \$	24488.801		
%POP <18 YRS	23.060		
%CMMT-AUTO	73.120		
%HU 3+ UNIT	8.400		
%HH FEMALE H	18.860		
%HU >1960	7.370	
%MOVED-1965	41.740	
%POP <5 YRS	6.620	
%MOVED<1960	41.890	
MEAN AGE	39.910	
%POP >65 YRS.	15.710	
%HU <1940	42.730	
%BLACK	1.850	
INC<125%POOR.	0.050	
%SERVICES	8.000	
%INC<\$5000	7.090	
%WHITE	93.900	
%OF MIXED	21.720	
%FEMALE>14LF.	41.560	
EMPLOY/POP	0.450	
%CLERIC+SALE.	32.970	
%CHICANO	2.450	
%DU>1.01/RM	1.870	
INC<125%POOR.	0.050	
%INC<\$5000	7.090	
%POP <5 YRS	6.620	
%>25 W HS	61.650	
MEAN SCHL YR.	11.400	
%CHINESE	3.060	
%OTHER RACES.	4.250	
%OF NATIVE	69.330	
PERSON/HH	2.830	
%GROUP Q	4.940	

CENSUS TRACT #408000.

OAKLAND

%HU 1 UNIT	98.540		
%OWN OCCPD	91.540		
HOME \$/INC	1.960		
%2+CARS	74.090		
AVG FAMILY S	2.970		
%INC>\$15000	61.130		
MEAN HOME \$	38698.801		
%POP <18 YRS	24.950		
%CMMT-AUTO	74.520		
%HU 3+ UNIT	0.660		
%HH FEMALE H	11.760		
%HU >1960	23.950	
%MOVED-1965	21.770	
%POP <5 YRS	4.460	
%MOVED<1960	26.990	
MEAN AGE	33.920	
%POP >65 YRS.	6.750	
%HU <1940	9.610	
%BLACK	2.610	
INC<125%POOR.	0.050	
%SERVICES	9.460	
%INC<\$5000	6.980	
%WHITE	90.790	
%OF MIXED	22.830	
%FEMALE>14LF.	45.870	
EMPLOY/POP	0.460	
%CLERIC+SALE.	30.990	
%CHICANO	2.830	
%DU>1.01/RM	0.0	
INC<125%POOR.	0.050	
%INC<\$5000	6.980	
%POP <5 YRS	4.460	
%>25 W HS	88.450	
MEAN SCHL YR.	13.700	
%CHINESE	4.460	
%OTHER RACES.	6.600	
%OF NATIVE	66.550	
PERSON/HH	3.720	
%GROUP Q	16.360	

%HU 1 UNIT	81.970		
%OWN OCCPD	78.160		
HOME \$/INC	1.670		
%2+CARS	70.030		
AVG FAMILY S	3.070		
%INC>\$15000	58.380		
MEAN HOME \$	42748.602		
%POP <18 YRS	31.210		
%CMMT-AUTO	88.360		
%HU 3+ UNIT	15.550		
%HH FEMALE H	11.520		
%HU >1960	62.030		
%MOVED-1965	69.140		
%POP <5 YRS	8.720		
%MOVED<1960	17.390		
MEAN AGE	32.170		
%POP >65 YRS.	5.030		
%HU <1940	5.650		
%BLACK	3.500		
INC<125%POOR.	0.030		
%SERVICES	5.700		
%INC<\$5000	2.620		
%WHITE	88.260		
%OF MIXED	25.190		
%FEMALE>14LF.	41.760		
EMPLOY/POP	0.450		
%CLERIC+SALE.	31.940		
%CHICANO	8.600		
%DU>1.01/RM	0.740		
INC<125%POOR.	0.030		
%INC<\$5000	2.620		
%POP <5 YRS	8.720		
%>25 W HS	87.000		
MEAN SCHL YR.	13.400		
%CHINESE	6.120		
%OTHER RACES.	8.240		
%OF NATIVE	64.810		
PERSON/HH	3.240		
%GROUP Q	0.0		

OAKLAND

152

CENSUS TRACT #410000.

OAKLAND

%HU 1 UNIT	97.140		
%OWN OCCPD	91.410		
HOME \$/INC	2.070		
%2+CARS	64.350		
AVG FAMILY S	3.010		
%INC>\$15000	45.840		
MEAN HOME \$	36520.398		
%POP <18 YRS	29.170		
%CMMT-AUTO	82.680		
%HU 3+ UNIT	2.870		
%HH FEMALE H	10.560		
%HU >1960	33.430	
%MOVED-1965	38.670	
%POP <5 YRS	7.350	
%MOVED<1960	30.790	
MEAN AGE	35.290	
%POP >65 YRS.	7.820	
%HU <1940	9.170	
%BLACK	20.280	
INC<125%POOR.	0.020	
%SERVICES	7.370	
%INC<\$5000	3.490	
%WHITE	76.450	
%OF MIXED	20.710	
%FEMALE>14LF.	44.710	
EMPLOY/POP	0.450	
%CLERIC+SALE.	26.610	
%CHICANO	9.370	
%DU>1.01/RM	0.480	
INC<125%POOR.	0.020	
%INC<\$5000	3.490	
%POP <5 YRS	7.350	
%>25 W HS	73.730	
MEAN SCHL YR.	12.700	
%CHINESE	1.860	
%OTHER RACES.	3.280	
%OF NATIVE	75.550	
PERSON/HH		3.130
%GROUP Q		0.0

CENSUS TRACT #432800.

SAN LEANDRO

%HU 1 UNIT	90.290		
%OWN OCCPD	78.390		
HOME \$/INC	1.800		
%2+CARS	68.700		
AVG FAMILY S	3.050		
%INC>\$15000	47.760		
MEAN HOME \$	42638.602		
%POP <18 YRS	29.330		
%CMT-AUTO	88.790		
%HU 3+ UNIT	4.000		
%HH FEMALE H	9.110		
%HU >1960	36.340	
%MOVED-1965	46.790	
%POP <5 YRS	7.670	
%MOVED<1960	26.300	
MEAN AGE	33.420	
%POP >65 YRS.	6.070	
%HU <1940	7.050	
%BLACK	0.490	
INC<125%POOR.	0.040	
%SERVICES	7.640	
%INC<\$5000	8.350	
%WHITE	95.730	
%OF MIXED	23.570	
%FEMALE>14LF.	39.950	
EMPLOY/POP	0.430	
%CLERIC+SALE.	30.770	
%CHICANO		13.430
%DU>1.01/RM		2.670
INC<125%POOR.		0.040
%INC<\$5000		8.350
%POP <5 YRS		7.670
%>25 W HS		73.410
MEAN SCHL YR.		12.400
%CHINESE		1.370
%OTHER RACES.		3.780
%OF NATIVE		71.480
PERSON/HH		
%GROUP Q		
			3.150
			0.0

REMAINDER OF MCD (OR CCD)

155

5. SMALL AREA POPULATION ALLOCATIONS

The homogeneous nature of the social areas is the major consideration in the allocation of population by age and sex. The set of rationally defined subareas is the basis of a successful small area population projection method. Since the social areas are homogenous with respect to lifestyle, stage-in-life cycle, ethnic composition, and other factors, the age/sex composition of the member census tracts are similar. The social areas make the allocation efficient and accurate. In addition the nature of the social areas makes interpretation for health planning easy.

The general change in the regional population age structure, computed by the regional demographic model, is a secondary allocation consideration. A social area's age/sex structure is assumed to be relatively stable through the 15 year projection period. Any changes in this structure result from the consideration of the region's shift in age/sex structure. For example, demographers agree that the population is aging. There is a smaller percentage of total population in the under 21 age groups due to declining fertility and the aging of the post World War II baby boom. These and other influences must be considered to incorporate changes to the social areas' age/sex structures.

The Allocation Process

The data requirements begin with subarea (social area) base year population totals. ABAG's land use modeling system produces transportation zone allocations of regional population in five year cycles. A first step in the allocation process is the aggregation of transportation zone totals to social areas. For 1970, the base year, census tract data is summed to social areas. For 1975, some special census totals and estimates are available. The land use modeling system provides the 1980 and 1985 zonal population totals. For each of the years, 1970, 1975, 1980 and 1985 the allocation requires the (96) subarea population totals. One alternative to this projection procedure would be county and city planning department estimates and projections. Other information required is regional population totals for 1970, 1975, 1980 and 1985 by age and sex. The regional demographic model, APPLE, produces the necessary tables.

The base year age/sex structure of each social area must be calculated. For health planning purposes fourteen age categories are needed: (1) 0 - 4, (2) 5 - 9, (3) 10 - 14, (4) 15 - 19, (5) 20 - 24, (6) 25 - 29, (7) 30 - 34, (8) 35 - 39, (9) 40 - 44, (10) 45 - 49, (11) 50 - 54, (12) 55 - 59, (13) 60 - 64, (14) 65 and over, for males and females. Twenty-eight cohort proportions of total social area population are calculated for each of the 96 areas using 1970 census tract data.

The last requirement, that of expressing a region-wide effect, can be solved in a number of ways. Using APPLE's regional projections of age and sex, cohort change factors were developed. These factors express at a regional level the change a particular cohort experiences from cycle to cycle. The factors are the ratios of cohort i in time period $t + 1$ to cohort i in time period t .

Process Flow

The Areal Demographic Allocation Model (ADAM) allocates the regional population total in five-year cycles. The cycle begins with the 1970 age/sex structure of each social area. For non-base year cycles this structure is adjusted by the regional change factors by multiplying each of the 28 cohort proportions by the appropriate cohort change ratio. For example, if the cohort males 10-14 increased regionally by six percent during a five-year period the change factor for that cohort is 1.06. The proportion of each social areas population for that age/sex cohort is multiplied by 1.06. Table 33 is an example of the cohort adjustment for social area 58.

The second step in each cycle is the multiplication of the area's population total by the adjusted cohort proportions. Since these proportions no longer sum to unity, the application of these adjustments produces a total population that is slightly different from the input social area population total.

The final step controls the projected cohort subtotals to the social area input totals and corrects for round-off errors. After the 28 age/sex cohorts are calculated for each social area the subarea total is compared to the input total. This row-wise checking is done to maintain structural compatibility between ADAM and input controls. If the totals differ for a social area the cohorts are adjusted equally. Once all 96 social area totals have been allocated the totals for each of the 28 cohorts (summed across all areas) are compared to the regional demographic model control totals. The comparison of these totals to the regional cohort control totals is printed. Adjustments must be made by the demographer or health planner who knows the region. The results of the factor analysis serves as a guide in deciding which areas may be adjusted. For example, if the older cohorts are underprojected, the social areas that are "old" on the stage in life cycle dimension, are the areas to adjust. Table 34 is an example of the social area projections for subarea 58. The model begins with 1970 base year data and projected 1975, 1980, and 1985.

Details on using and preparing data for ADAM are given in the software documentation volume.

Allocation Refinements

The 96 social areas of the nine county region are very diverse, and provide significant challenges to the allocation process. In creating the social areas, small, very different areas emerged that possessed such distinct characteristics that they would not be annexed by their neighbors. These isolates or special areas pose major problems to the allocation scheme. Such social areas are characterized by either a skewed age/sex distribution, a high group quarters population, or perhaps a military base or a university campus. These social areas usually have small populations. A threshold of 6,000 total population was chosen to define areas deserving special attention. To complicate the process further, the land use model's projections of populations for these already small areas usually indicate declines, beginning in 1975. For these areas,

TABLE 33
SAMPLE SOCIAL AREA AGE/SEX STRUCTURE

ADJUSTED AGE AND SEX DISTRIBUTION

SOCIAL AREA 58
1970

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	0.0245	0.0321	0.0398	0.0479	0.0372	0.0302	0.0276	0.0233	0.0237	0.0352	0.0354	0.0328	0.0292	0.0509	0.4748
FEMALES	0.0314	0.0305	0.0370	0.0499	0.0470	0.0313	0.0250	0.0216	0.0293	0.0402	0.0403	0.0368	0.0322	0.0741	0.5266
TOTAL	0.0609	0.0626	0.0768	0.0978	0.0842	0.0615	0.0526	0.0449	0.0530	0.0754	0.0757	0.0696	0.0614	0.1250	1.0014

SOCIAL AREA 58
1975

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	0.0253	0.0251	0.0366	0.0486	0.0377	0.0294	0.0315	0.0235	0.0208	0.0306	0.0368	0.0348	0.0313	0.0807	0.4921
FEMALES	0.0264	0.0235	0.0323	0.0480	0.0412	0.0337	0.0286	0.0214	0.0240	0.0342	0.0417	0.0378	0.0335	0.0830	0.5093
TOTAL	0.0517	0.0486	0.0683	0.0966	0.0789	0.0631	0.0601	0.0449	0.0448	0.0648	0.0785	0.0726	0.0648	0.1637	1.0014

SOCIAL AREA 58
1980

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	0.0261	0.0230	0.0307	0.0474	0.0411	0.0323	0.0328	0.0265	0.0230	0.0288	0.0333	0.0374	0.0335	0.0630	0.4789
FEMALES	0.0275	0.0215	0.0279	0.0467	0.0446	0.0326	0.0337	0.0265	0.0265	0.0307	0.0382	0.0418	0.0353	0.0889	0.5224
TOTAL	0.0536	0.0445	0.0586	0.0941	0.0857	0.0649	0.0665	0.0530	0.0495	0.0595	0.0715	0.0792	0.0688	0.1519	1.0013

SOCIAL AREA 58
1985

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	0.0287	0.0254	0.0303	0.0435	0.0430	0.0379	0.0386	0.0297	0.0297	0.0341	0.0337	0.0364	0.0388	0.0721	0.5219
FEMALES	0.0302	0.0241	0.0275	0.0435	0.0467	0.0378	0.0350	0.0335	0.0350	0.0362	0.0367	0.0411	0.0419	0.0103	0.4795
TOTAL	0.0589	0.0495	0.0578	0.0870	0.0897	0.0757	0.0736	0.0632	0.0647	0.0703	0.0704	0.0775	0.0807	0.0824	1.0014

TABLE 34

SAMPLE SOCIAL AREA POPULATION PROJECTION

PROPOSED AGE AND SEX DISTRIBUTION

SOCIAL AREA 58 1970															
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	1473.	1603.	1988.	2392.	1858.	1508.	1378.	1163.	1183.	1758.	1768.	1638.	1458.	2542.	23710.
FEMALES	1568.	1523.	1848.	2492.	2347.	1563.	1248.	1078.	1463.	2008.	2013.	1838.	1608.	3702.	26299.
TOTAL	3041.	3126.	3836.	4884.	4205.	3071.	2626.	2241.	2646.	3766.	3781.	3476.	3066.	6244.	50009.
SOCIAL AREA 58 1975															
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	1199.	1188.	1706.	2301.	1788.	1392.	1490.	1113.	985.	1449.	1742.	1646.	1481.	3827.	23307.
FEMALES	1250.	1113.	1529.	2275.	1952.	1597.	1352.	1014.	1137.	1622.	1975.	1791.	1587.	3934.	24128.
TOTAL	2449.	2301.	3235.	4576.	3740.	2989.	2842.	2127.	2122.	3071.	3717.	3437.	3068.	7761.	47435.
SOCIAL AREA 58 1980															
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	1458.	1282.	1712.	2645.	2292.	1802.	1830.	1480.	1283.	1604.	1859.	2086.	1871.	3517.	26721.
FEMALES	1535.	1201.	1554.	2609.	2487.	1817.	1881.	1480.	1476.	1710.	2131.	2333.	1971.	4965.	29150.
TOTAL	2993.	2483.	3266.	5254.	4779.	3619.	3711.	2960.	2759.	3314.	3990.	4419.	3842.	8482.	55871.
SOCIAL AREA 58 1985															
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	TOTAL
MALES	1741.	1543.	1836.	2642.	2611.	2300.	2344.	1801.	1804.	2070.	2045.	2210.	2356.	4381.	31684.
FEMALES	1832.	1460.	1668.	2639.	2836.	2292.	2127.	2033.	2127.	2196.	2231.	2498.	2546.	622.	29107.
TOTAL	3573.	3003.	3504.	5281.	5447.	4592.	4471.	3834.	3931.	4266.	4276.	4708.	4902.	5003.	60791.

analysts depended heavily on the theory and results of the factor analysis. These areas' social characteristics are more likely to remain constant. Especially high factor scores are evident in the age/sex related factors for these special areas.

The special areas' age and sex distributions are held constant and applied to the land use model's control total. For any institutional situation, therefore, the age/sex structure is constant.

Results and Validation

A sample result of the allocation process is Table 34. For this social area in Alameda County, population increases steadily. With its middle class suburb characteristics, this area responds well to the process. A decline in fertility as well as the impacts of the region's relatively high immigration rate are evident in this disaggregation.

The social areas covered by a 1975 census were studied closely. The 1975 population by age and sex represented the new census data source. That the 1975 census age and sex distribution and the ADAM results are very similar for the five census counties is not surprising. The counties taking censuses in 1975 can be characterized as areas of steady suburban growth. The social areas within these counties are in the early part of the stage in family life cycle. These cohorts are easily described by the existing model.

Strict control was not maintained to regional cohort totals. Each of the 28 cohorts was added across social areas to obtain a regional cohort total. When this total is compared to the regional demographic model's cohort total a bias was revealed. The factors used to represent regional demographic trends are less effective in the 65+ age group than in any other cohort. In 1975 there is less than 1% under-representation for this age group. By 1985 the error has grown to just less than 3% under-representation. The other cohorts are modeled more successfully. Demographers recognize an unprecedented increase in the ranks of the over 65. From the bias described above it appears that ADAM in its present design reacts too conservatively to the increases in this age group. Overall, the social area projections are successful.

6. SUMMARY AND SHORTCUTS

The success of the small area population projection methodology package discussed above is due to the combination of the regional concept and rationally defined subareas. The package and examples of the second section are the maximum effort. A few shortcuts are mentioned in the text. The minimum effort must clearly include: (1) correct definition of the region, (2) grouping analysis with key social variables to define the subareas. The social area approach makes the models more than a mechanical allocation. The approach provides the interpretive framework, linking the tables of projected population to the subarea descriptions. The result is a useful means for the health planner to understand the nature of the population as it affects health resource use.

U.C. BERKELEY LIBRARIES



C124909696